

City University of New York (CUNY)

CUNY Academic Works

Dissertations, Theses, and Capstone Projects

CUNY Graduate Center

9-2021

Climbing The Mountain When There Is No Mountain To Climb: Pragmatism and the Reconstruction of Moral Philosophy

Ryan Marshall Felder

The Graduate Center, City University of New York

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/gc_etds/4500

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).

Contact: AcademicWorks@cuny.edu

CLIMBING THE MOUNTAIN WHEN THERE IS NO MOUNTAIN TO CLIMB:
PRAGMATISM AND THE RECONSTRUCTION OF MORAL PHILOSOPHY

by

RYAN MARSHALL FELDER

A Dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the
requirements for the degree of Doctor of Philosophy, The City University of New York

2021

© 2021

RYAN MARSHALL FELDER

All Rights Reserved

Climbing The Mountain When There Is No Mountain To Climb: Pragmatism and the
Reconstruction of Moral Philosophy

by

Ryan Marshall Felder

This manuscript has been read and accepted for the Graduate Faculty in
Philosophy in satisfaction of the dissertation requirement for the degree of
Doctor of Philosophy.

Date

Miranda Fricker

Chair of Examining Committee

Date

Nickolas Pappas

Executive Officer

Supervisory Committee:

Jesse Prinz

Miranda Fricker

Gary Ostertag

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Climbing The Mountain When There Is No Mountain To Climb: Pragmatism and the

Reconstruction of Moral Philosophy

by

Ryan Marshall Felder

Advisor: Jesse Prinz

The aim of this dissertation is to leverage philosophical resources given to us by Richard Rorty which can show that various diverse and seemingly-incompatible strands in moral theory can be reconciled to show that the differences between moral theories are illusory. The strategy is to show that Rorty's approach to philosophy allows a pragmatist reconstruction of several prominent positions in normative ethics (Chapters 1 and 2) and metaethics (Chapter 3 and 4), and that a Rorty-style pragmatism, with its emphasis on liberal ironism and a relaxed approach to semantics and truth, facilitates the moderation of normative theories and metaethical theories, and hence that a Rortian perspective allows discourse between diverse and seemingly-incompatible positions. This result is interesting for two reasons. First, it undermines pragmatist orthodoxy, encouraged by Rorty himself, that pragmatists should not try to construct theories. I show that this view ignores the possibility that moral and metaethical theorizing may be aided by the acceptance of a Rortian perspective, and that methodology in ethics in important ways presupposes the truth of pragmatism. Second, the project provides interesting answers to ethical questions along the way, and hence indirectly shows that pragmatists can productively engage in technical philosophical debates and make legitimate positive contributions. Hence, I see myself as making an indirect argument for the Rortian perspective.

ACKNOWLEDGEMENTS

I owe significant debts to several people for support and assistance with this project. First, I want to thank my committee, Miranda Fricker, Gary Ostertag, and Jesse Prinz, for the academic support through the prospectus and dissertation writing process that made this finished product possible. In particular, I want to thank Jesse for reading through every page multiple times, with excruciating attention to detail, and for being tolerant of the project's numerous unresolved issues. All intellectual flaws are, of course, my own. I also owe thanks to Steven Ross, who oversaw my Qualifying Paper 2 writing process, a paper which eventually turned into Chapter 3 of this dissertation. Finally, I owe a huge debt to Rosamond Rhodes, who through her generous advising at Mount Sinai opened up the professional pathway that made my dissertation defense possible. Without the opportunities and professional development I received while working under her, this document truly could not exist.

I would like to thank Gary Santillanes, who more than anyone else taught me how to be a philosopher, and to understand the value that I can bring to an academic institution.

I also must thank my father, David Felder, for a lifetime of love and support, and for teaching me to divide by zero.

Finally, and most importantly, this dissertation is dedicated to my wife, Stella Morrison. Thanks for putting up with a decade of this nonsense.

TABLE OF CONTENTS

PREFACE: A Note on the Plan of the Work	1
INTRODUCTION: Multi-Level Moral Theories and the Liberal Ironist	3
CHAPTER 1: Esoteric Morality: A Pragmatist Reconstruction of Some Aspects of Motive-Consequentialism	12
1.1. Introduction: plan of the chapter	12
1.2. Multi-level consequentialism and the unavoidability of esoteric morality	12
1.3. Rortian neopragmatism and the esoteric liberal ironist	20
1.4. The esoteric character of conceptual feminism	29
1.5. Weaving together the threads	38
CHAPTER 2: Reconstructing Moderate Deontology in Political Liberalism	40
2.1. Introduction: moderate and threshold deontologies	40
2.2. The Legal Model and Rawls's four-stage sequence	44
2.3. Mid-level principles	50
2.4. Non-lexical moderate deontology: answering Smilansky's challenge	53
2.5. Rorty, political liberalism, and the method of moderate deontology	57
INTERLUDE: From Pragmatist Normative Ethics to Pragmatist Metaethics	63
CHAPTER 3: Intuitions in Ethics From a Pragmatist Perspective	65
3.1. Introduction: pragmatist skepticism about intuitions	65
3.2. Some views of intuitions	67
3.3. A methodological problem with the empiricist conception of intuitions	77
3.4. Methodological problems with the rationalist conception of intuitions	79
3.5. Pragmatism and the natural kind theory of intuitions	86

CHAPTER 4: A Pragmatist Metaethics? Reconstructing the Terrain	98
4.1. Introduction: plan of the chapter	98
4.2. Non-cognitivism	98
4.3. Error theory	107
4.4. Cornell realism	111
4.5. Constructivisms Humean and Kantian	120
CONCLUSION: Pragmatism, Reconstruction, and Philosophy on the Cusp of the Real World	136
BIBLIOGRAPHY	138

PREFACE: A Note on the Plan of the Work

The purpose of this dissertation is to explore what it is like, in a broad sense, for a Rortian pragmatist to accept and defend an ethical view. As I will argue in the Introduction, ethical views, and moral theories in particular, present a troubling tension for pragmatists. Part 1 of this dissertation, comprising Chapters 1 and 2, focuses on this problem from the standpoint of normative ethics, here understood as the aim of constructing a moral theory that provides principles for dealing with ethical hard cases. Yet this project raises more conceptual, theoretical questions about whether any sense can be made of the underlying methodological, semantic, and epistemological assumptions thereby made. Part 2, comprising Chapters 3 and 4, hence takes on the results of Part 1 from the standpoint of metaethics, and attempts to smooth over possible metaethical objections to the purported links between normative ethics and pragmatism.

The Introduction to this work only addresses the material covered in Part 1. For an Introduction to Part 2, the reader should wait until the Interlude, which motivates the reasons for pursuing the metaethical foundations of the claims made in Part 1. The Conclusion wraps the whole thing up by connecting the results of Parts 1 and 2 to the practice of applied and professional ethics.

In my view, Parts 1 and 2 are deeply intertwined, and best read as a whole, because normative ethics raises questions about the possibility of, for instance, moral knowledge, which is addressed by a long tradition and a voluminous literature, and correspondingly because metaethics demands application as an underlying buttress for normative ethics. To paraphrase Kant's famous dictum (1998, A51/B75), I believe that normative ethics without metaethics is empty, and metaethics without normative ethics is blind. Nonetheless, there is also a tradition of treating normative ethics and metaethics as separate topics. Therefore, those inclined to do so can

read Part 1 and Part 2 separately from each other: Part 1 as an excursion in normative ethics from the pragmatist perspective, Part 2 as an attempt to provide some philosophical defenses of the pragmatist perspective itself.

INTRODUCTION: Multi-Level Moral Theories and the Liberal Ironist

The purpose of this dissertation is to explore what it is like for a Rortian pragmatist to accept and defend an ethical view. Although there is nothing in Rorty that tells us not to hold ethical views, there are features of ethical views that can be hard to capture if one accepts his pragmatist, ironist, theory-opposed metaphilosophical framework. For instance, most ethicists would agree that ethical views, insofar as they attempt to provide answers to practical questions, aim at providing the final word on their subject. A moral theorist constructs a moral theory because they want to provide the best theoretical explanation of the moral phenomena. Rorty, however, insisted that there are no final vocabularies, ethical or otherwise, that can foreclose inquiry into those vocabularies' foundations (Rorty 1979, pp. 206ff). Hence, Rorty sat uncomfortably, at best, with moral theories. As evidence of this, Rorty encapsulated this attitude in the Introduction to *Contingency, Irony, and Solidarity* (Rorty 1989, p. xv) as follows:

For liberal ironists, there is no answer to the question “Why not be cruel?” – no non-circular theoretical backup for the belief that cruelty is horrible. [...] This question strikes liberal ironists as just as hopeless as the questions “Is it right to deliver innocents over to be tortured to save the lives of $m \times n$ other innocents? If so, what are the correct values of n and m ?” [...] Anybody who thinks that there are well-grounded theoretical answers to this sort of question – algorithms for resolving moral dilemmas of this sort – is still, in his heart, a theologian or a metaphysician. He believes in an order beyond time and change which both determines the point of human existence and establishes a hierarchy of responsibilities.

One reason for this is that all vocabularies are historically- and socially-contingent: we pick up the beliefs we have at least in part through socialization, and the aim of providing the final word on a subject is to deny this historicist nature of our vocabularies (Rorty 1979, pp. 9-10). Hence, as a matter of boring empirical reality, vocabularies tend to reproduce themselves over time and sustain themselves through changes, due to the dialectical conflict between the historicist nature of vocabularies and the fact that vocabularies are not homogenous within any group (Rorty 1989,

pp. 8-9). But a moral theory, assuming that it tries to provide the best explanation, is not historically- and socially-contingent, except in the uninteresting sense in which new principles and rules must be proposed as history advances. Hence the discomfort implicit in Rorty's view of traditional normative ethics: vocabularies are historically- and socially-contingent, and moral theories can be no such thing. However, as we will see, there is some reason to think that this discomfort can be dissolved.

Before this can be explained, though, I have used a number of terms without explaining them. I want to rectify this by introducing definitions for 'vocabulary', 'pragmatism', and 'liberal ironist'. First, pragmatism. The nature and aims of pragmatism are the subject of disagreement even among deeply-committed pragmatists. Rather than being comprehensive, I invoke William James's pragmatist notion of the aims of ethics in his oft-quoted paper "The Moral Philosophy and the Moral Life":

The main purposes of this paper is to show that there is no such thing possible as an ethical philosophy dogmatically made up in advance. We all help to determine the content of ethical philosophy so far as we contribute to the race's moral life. In other words, there can be no final Truth in Ethics any more than in Physics, until the last man has had his experience and said his say (James 1891, p. 330; see also Moody-Adams 1997, p. 153).

Michele Moody-Adams, whose pragmatist-leaning discussion of moral realism in *Fieldwork in Familiar Places* is brought up for consideration in Chapter 4, approvingly cites James here. Although James writes in a way that suggests that a final ethical vocabulary is possible, it is important to keep in mind that James's use of the term 'last man' implies that we will never achieve a final ethical vocabulary, for no particular strand of ethical inquiry will be expansive enough in scope to meet that criterion. The best we can hope for is to simply extend ethical discourse to as many people as possible and to encompass the broadest array of ethical concerns. But of course, this means that we will simply stop asking questions like the questions quoted in

the *Contingency, Irony, and Solidary* passage above. To be a pragmatist in my sense hence means to aim at dissolving traditional philosophical problematics, such as the problematic about which values of n and m are correct (Rorty 1991c, p. 127-8). Pragmatism also allows us to think as philosophical theories as tools, rather than as attempts to map the reality in their domains. We care about ethical vocabularies because they give us a standard by which to approach real problems, and hence because they allow people to “cope with their environment” (Rorty 1982a, p. xviii; see also West 1989, p. 98).

A crucial point that this dissertation defends is that, even given this idea of pragmatism, there is no reason to think that people need to abandon their ethical vocabularies if they accept pragmatism. Instead, people have to balance their commitment to their ethical vocabularies with the metaphilosophical pressure that pragmatism imposes on them. Rorty thinks that this shows that individuals in liberal societies, for whom the pragmatist pressures can be significant, are pushed to view their ethical vocabularies ironically. Hence, Rorty’s character of the liberal ironist: one who is faced with the continual challenge of, on the one hand, accepting the contingency of the vocabulary that they endorse, and on the other, actually maintaining genuine commitment to that vocabulary (Rorty 1989, p. 61).

Finally, a bit on what I will mean by a vocabulary. This way of talking rises out of Quine’s attacks on the verification theory of meaning, an attack whose consequence is that holism about meaning is the best available account of meaning. The basic idea is that sentence meanings are determined by a sentence’s relation to various other sentences, and not individually (Quine 1951, pp. 39-40). Rorty, following Quine, targets whole vocabularies and their collective meanings as the proper object of analysis (Rorty 1979, pp. 193-5). We produce lots of kinds of vocabularies, each of which has some purpose. We have, for instance, vocabularies of scientific

inquiry. We also have vocabularies for ethics. But sometimes, it is argued that a vocabulary is final. This implies that the vocabulary gives the final word on a domain. If there were an ethical final vocabulary, then that vocabulary would state the final word on all cases in ethics, always offering up the right prescription about what we ought to do. In other words, a final ethical vocabulary would be a gold-standard moral theory. But as I have already said, Rorty thinks that no vocabulary can be final. One reason is the historical and social contingency of vocabularies. Another reason, which I discuss in more detail below, is that even a vocabulary that states the final word on all cases cannot be taken to be the only way to respond to that case. A vocabulary about domain D, hence, is a group of meanings, concepts, and presuppositions that outline the things that we can say about D. A final vocabulary purports to exclusively and exhaustively explain D.

The alleged problem for being able to accept, use, and advocate for a moral theory arises because of the distinctive pragmatist standpoint that Rorty views as necessary. Yet the problem is not quite as bad as we might think. An ethical vocabulary can contain as one of its constituent parts a moral theory, and people who accept different ethical vocabularies often as a consequence of this accept different moral theories. Hence, even on pragmatist assumptions that make it impossible to say that a theory is the true, final theory, we cannot think that a liberal ironist is indifferent toward their vocabulary. If a liberal ironist is to make sense of their own commitments, they need a way to understand these commitments as integrated in some way. Rorty's answer to this question is to hang a lot of weight on the public/private distinction, and to exile questions about ethical vocabularies to the private side, leaving public morality to be conducted in the Rawlsian vocabulary of political liberalism (Rorty 1991d). Rorty also thinks

that works of literature, which sensitize us to various kinds of cruelty that occur in liberal societies, can help us to probe the outer limits of our ethical vocabularies (Rorty 1989, p. 53).

The aim of the next two chapters, and to a lesser extent the whole dissertation, is to test these claims. Is Rorty correct to say that, given the private/public distinction, the most we can hope for, in terms of giving some philosophical structure to moral life, is allowing the ‘private’, philosophical side to accept all moral theory as an intellectual exile, while letting public morality be solely constituted by the terms of political liberalism? I believe that he is not. The private/public distinction is itself a matter of philosophical and political adjudication, and private ethical vocabularies can, in virtue of their internal substance, be modeled after the public to various degrees. When stated out loud, this should not be surprising. Feminists have long argued that the various forms and uses of the private/public distinction are tools for instituting patriarchy (Okin 1991, ch. 6). This should at least establish the possibility that this distinction is malleable, or even rejectable, for some good pragmatist reasons, because Rorty views himself as closely allied with feminist aims (Rorty 1998b). Given the pragmatist aims of attacking the presuppositions of philosophy, this distinction needs to be on the table, but Rorty’s discussions in *Contingency, Irony, and Solidarity* suggest that it shouldn’t be. In Chapters 1 and 2, I defend these claims by providing a proof of concept that attacking the distinction is prescribed by the pragmatist framework and the position of the liberal ironist. We should reject the private/public distinction because allowing private moral theories to be reconstructed in light of public democratic values that supposedly establish the possibility of a liberal society would be a good thing, on pragmatist grounds. It offers both the possibility of an inclusive public morality and an interlocking system of private moral theories that both support and are supported by public

morality, and hence the possibility of overcoming the discomfort expressed by Rorty toward normative ethics in *Contingency, Irony, and Solidarity*.

Some further discussion of my sense of pragmatism would be helpful for explaining how a liberal ironist can have a moral theory. In using the term reconstruction, which appears in the dissertation's title, to explain my project, I am making a nod to Dewey's project in *Reconstruction in Philosophy*. There, he writes of a traditional philosophical conception of rationality as "a power transcending in origin and content any and all conceivable experience can attain to universal, necessary and certain authority and direction" (Dewey 1920, p. 78; see also Dewey 2018, p. 314). Yet this conception of rationality, Dewey argues, needs to be rejected in favor of a more modest conception of rationality that brings reasoning closer to people's experiences of reasoning (Dewey 1920, p. 87). The pattern of argument: rationality is a practice of achieving knowledge, and we need practices of achieving knowledge because of the general characteristics of our experiences. Rationality is a way to cope with experience, as is philosophy for a Rortian pragmatist. Extrapolating quite a bit, we can say that reconstruction is a way to strategically modify our philosophical conceptions in light of the failures of traditional conceptions (Dewey 1920, pp. 95-96). Hence, when the conception of rationality that we accept doesn't help to cope with experience, we should reject it and provide a replacement theory that better answers to the demands that the practical world, broadly construed, places on it. From this arises the impulse to provide what I call multi-level moral theories. A multi-level moral theory is distinguished from a single-level moral theory in having different internal aspects that issue in their own dictates and potentially point in different directions. A single-level moral theory, by contrast, has only one internal aspect that issues in dictates. Peter Singer's conception of utilitarianism in "Famine, Affluence, and Morality" is an example of a single-level moral theory

because it endorses only the principle of utility as a decision criterion for settling questions about our obligations to the global poor. Singer explicitly denies attempts to turn his utilitarianism into a multi-level moral theory, when he argues against Sidgwick's views on moral codes (Singer 1971, pp. 237-8). On such a view, utilitarianism is not single-level but has multiple internal levels, according to which some people ought to believe utilitarianism and act according to it, and other people ought not to believe utilitarianism and instead ought to act on some other, different theory.

Rule consequentialism, which is what Singer rejects, and moderate deontology are two important examples of multi-level moral theories. In elaborating these theories and giving them indirect defenses in Chapters 1 and 2, I aim to show that elaborating a multi-level moral theory is aided by thinking of moral theory as a kind of reconstruction of our basic ethical ideas. The thesis of this part of my dissertation, therefore, is that multi-level moral theories are a good philosophical tool available for a pragmatist liberal ironist to use when confronting ethical challenges, public or private, and to properly accommodate the inevitability of the contingency of our vocabularies.

Part of what makes this project interesting is that I pursue distinctively pragmatist attempts at reconstructions. But pragmatism and Richard Rorty are both controversial. Hence, why bother to draw on them so heavily? The reason is that not all reconstructions are pragmatist; philosophical realists have also attempted to reconstruct moral theories. As we'll see, pragmatism typically provides a more satisfying set of reconstructed moral theories than does realism. For example, in Chapter 1, I discuss Derek Parfit's attempts to reconstruct consequentialism in light of the considerations of common-sense morality, which is often taken to have strong anti-consequentialist content. In *Reasons and Persons*, Parfit seems himself as

arguing that the distance between consequentialism and common-sense morality is not as large as some might have thought, and that we can establish the truth of consequentialism by showing that it can be modified to accommodate insights forced upon it by common-sense morality:

[Consequentialism and Common-Sense Morality] face objections that can be met only by enlarging and revising those theories, in ways that bring them closer together. These facts natural suggest an attractive possibility. [Consequentialism and Common-Sense Morality] may dovetail, or join together to make a larger whole. We might be able to develop and develop a theory that includes and combines revised versions of both [...]. Call this the *Unified Theory* (Parfit 1984, p. 112).

Continuing this project in *On What Matters, Volume One*, Parfit (2011a, pp. 411ff) argues that consequentialism, contractualism, and deontology are all paths up the same ethical mountain, and that reconstruction in ethics is possible because disparate moral theories all culminate in the same moral truth. Although Parfit's goal is admirable, and partly congenial to the project of reconstruction I pursue here, it turns out that Parfit's insistence on realism about these reconstructions is problematic. For instance, although Parfit speaks of reconstruction, he still thinks of truth or some closely related notion of theory justification as the criterion for a successful theory (Parfit 1984, p. 24). This is problematic because, even granting that a reconstruction aims at truth, there might be some theory that helps bridge the gaps between competing theories, but is indifferent to the truth. In view of the multiple and diverging aims that a moral theory might be thought to serve, the Parfitian realist reconstruction would potentially ignore helpful reconstructions. I will argue, therefore, that insofar as reconstruction is crucial in view of the general concerns of moral life and the real-world considerations that moral theory responds to, and that pragmatism offers the more useful way of thinking about reconstructions, this offers some reason to doubt a realist attitude toward ethics. It turns out that we can climb Parfit's mountain, regardless of whether a mountain exists in the first place.

Pragmatism, as Rorty conceives it, tells us that “it is the vocabulary of practice rather than theory, of action rather than contemplation, in which one can say something useful about philosophical conceptions” (Rorty 1982b, p. 162). I aim to provide theories that are thoroughly couched in terms of practice. This conception of pragmatism stresses the importance of doing philosophy, rather than merely contemplating. Hence, the best way to elucidate the value of pragmatist reconstruction is by actually doing philosophy, in this case normative ethics. So now, the task is to do normative ethics as a pragmatist might. To normative ethics we turn.

CHAPTER 1: Esoteric Morality: A Pragmatist Reconstruction of Some Aspects of Motive-Consequentialism

1.1. Introduction: plan of the chapter

The aim of this chapter is to weave together three important moral ideas that, at first, appear to have very little to do to each other, yet on further inspection have deep structural similarities: multi-level consequentialism, Rortian neopragmatism, and conceptual feminism. By weaving them together, I offer an indirect defense of multi-level consequentialism, insofar as this weaving together shows that certain controversial implications of multi-level consequentialism are really just a feature of accepting a moral theory in a world like ours. Moral theories that attribute different decision procedures depending on the constitution of the person using the theory and the context in which it is applied are simply, contra contractualist worries, a feature of moral agents existing in a complex world. To see that this broad variety of positions can maintain an esoteric status while still providing a moral perspective worthy of our belief and allegiance is to see that attempts to reject those theories by appealing to a contractualist notion of publicity are fundamentally misguided. Hence, the indirect defense of multi-level consequentialism takes a distinctively Rortian, liberal ironist form. This defense, as we will see, has the added virtue of bring consequentialism and Rortianism into conversation with contemporary developments in feminist philosophy in a surprising and unexpected way.

1.2. Multi-level consequentialism and the unavoidability of esoteric morality

When discussing multi-level moral theories, perhaps the first thing to come to mind is rule-consequentialism. Rule-consequentialism is an attempt to reformulate the notion of consequentialism in response to some prominent objections. From the *Stanford Encyclopedia of Philosophy*, we can begin with Walter Sinnott-Armstrong's statement of these ideas:

consequentialism tells us that the rightness of an act “depends only on the consequences of that act or of something related to that act, such as the motive behind the act or a general rule requiring acts of the same kind” (Sinnott-Armstrong 2019). This claim states the basic idea of consequentialism, that when we are engaged in moral evaluation, we should look to the consequences of the thing being evaluated, as well as the basic idea of rule-consequentialism, which is that (at least) sometimes, the thing being evaluated is a rule for conduct. Given that the discussion that follows is largely framed in terms of Derek Parfit’s (1984) version of consequentialism, I focus on motive-consequentialism, which could also be called belief-consequentialism or desire-consequentialism. This version of consequentialism is also suggested by Sinnott-Armstrong’s definition: at least sometimes, the thing being evaluated by consequentialism is our motives for action.

Motives have a temporal element: having a motive to do something is often stable across time. This is important because it means that having motive M, which tells us to do A, might have different consequences than A’s just occurring. In particular, M might have consequences that affect our actions outside of the contexts where it tells us to do A. Consider, for instance, Parfit’s discussion of parent’s dilemma cases, in which a parent who is motivated to treat all children equally might bring about bad consequences over the course of their own child’s life, because having that motive leads them to neglect their special obligations to their own child (Parfit 1984, pp. 96-8). Although it might sometimes be best for a parent to treat all children equally in view of the consequences, a life-long motive to do that could lead that parent to impose other bad consequences on their child.

In a parent’s dilemma case, act-consequentialism and motive-consequentialism diverge; they tell us to do different things. Act-consequentialism would tell us to consider all children’s

well-being equally, but motive-consequentialism, insofar as it is temporally broader, would tell us to have motives that lead to our privileging our own child's well-being. Parfit explains the problem as follows:

It might be claimed that, if Consequentialism sometimes breaks the link between the belief that our act is wrong and the belief that we are bad, we would not in fact continue to regard morality with sufficient seriousness. Our desire to avoid wrongdoing may be undermined if we have other desires which are often stronger. This desire may survive only if we believe that it should *always* be overriding, and feel remorse when it is not. It might be claimed, on these or other grounds, that it would make the outcome better if we always keep the link between our moral beliefs and our intentions and emotions. If this is so, it would make the outcome better if we did not believe [Consequentialism] (Parfit 1984, p. 40).

It is because our motives are relevant to what we should do, in general, that the problem arises: if we have the motives that motive-consequentialism tells us to have in the parent's dilemma case, then motive-consequentialism is divided against act-consequentialism, because act-consequentialism doesn't look at the motives we have when determining how we should act. In 'breaking the link,' to borrow Parfit's phrase, between our actions and our motives, consequentialism presents us the problem of reconciling these various aspects of the theory.

Parfit takes himself to have a good answer to this problem, but others have viewed this as a knock-down argument against consequentialism itself. Bernard Williams, for instance, has connected this feature of this problem to what I will call the objection from esoteric morality: consequentialism tells the elite members of a society to keep the truth of consequentialism away from the common folk, and in virtue of that, we should reject consequentialism. The argument trades on the claim that it is better, from a consequentialist perspective, if the elites understand consequentialism as a philosophical theory when acting from it, while everyone else simply does the various things that consequentialism enjoins, justification-free (Williams 1987, p. 108). Anecdotally, it seems like consequentialism has this implication. In my own experiences of

teaching consequentialism and utilitarianism to my students, about half of them misunderstand those views as a kind of hedonism that simply tells us to pursue pleasure, despite repeated exposure to the distinction between utilitarianism and hedonism. The argument then claims that we should reject esoteric moral theories like Parfit's, because an esoteric moral theory cannot be a way of life for its adherents, and a moral theory must be able to provide a way of life for its adherents (Williams 1987, p. 110).

I want now to offer a more precise definition of an esoteric moral theory, to distinguish it from a related notion. According to Ben Eggleston (2013, p. 33), an esoteric moral theory T is one whose being believed directs *some* who believe in it to come to believe some other moral theory T*, and others to believe T. Eggleston notes that this notion is often confused with a self-effacing moral theory, which is a moral theory T whose being believed directs *anyone* who believes it to come to believe some other moral theory T*. In this chapter, I focus specifically on esoteric moral theories; I am more concerned with whether consequentialism tells, for instance, some parents in parent's dilemma cases to be motivated by common-sense morality, even granted the truth of consequentialism. This would be the case if there were two groups of parents in virtue of consequentialism in these cases: one elite group that considers the interests of all children equally, and another, larger, folk group that tries to discharge their special obligations toward their children. My reason for doing this is that the possibility of a genuinely self-effacing morality is hard to see. Parfit (1984, p. 41) writes, for instance, that it is very hard to see how the best consequences could be brought about if nobody, ever, anywhere ever tried to bring about the best consequences. But this is just what a self-effacing morality would require. Similarly, Katarzyna de Lazari-Radek and Peter Singer argue against EM by pointing out that, in a great number of cases, transparency in moral theorizing will bring about the best consequences, and so

it seems that at most C could be partly self-effacing (de Lazari-Radek and Singer 2010, pp. 51-4).

In view of these comments, it may be helpful to state the objection from esoteric morality as a syllogism:

1. Consequentialism is esoteric
2. We should reject a moral theory if it is esoteric
3. Therefore, we should reject consequentialism (Eggleston 2013, p. 38)

As I have already claimed, Parfit takes himself to have a good response to this objection. In Parfit's view, we should reject premise 2, because it supposes that it is the aim of a moral theory to be believed. The aim of a moral theory is not, however, to be believed, but to be true, or to be the right moral theory. A theory that meets that condition can be accepted, even if it sometimes tells some of its adherents to believe a different theory (Parfit 1984, p. 24). de Lazari-Radek and Singer (2010, p. 56) cite this claim uncritically in their defense of consequentialism against the objection from esoteric morality.

Although this response to the objection is plausible as a first pass, it raises the broader question of what constitutes the aim of moral theory. Parfit (1984, p. 43) rightly notes that this is not settled by his arguments in Part One of *Reasons and Persons*. But once this question has been opened up, it becomes plausible to wonder whether this conception of the aims of moral theory is correct after all. Although such a view is common among consequentialists, contractualist moral theories tell a very different story. On a contractualist moral theory, the aim of moral theory is to provide a set of moral principles that are the potential object of consensus among moral agents. T. M. Scanlon, for instance, thinks that a moral theory aims at the presentation of a set of moral principles none of which could be reasonably rejected by moral agents (Scanlon 1998, p. 189). Scanlon's discussions are an expansion and defense of similar

claims made by John Rawls, who argues that the aim of moral theory is to provide a publicly-acceptable set of moral principles which would be accepted by all suitably situated free and equal moral agents (Rawls 1999, pp. 10ff). This is the meat of Rawls's publicity constraint (Rawls 1999, p. 115).

The debate between those who accept an esoteric version of consequentialism and those who accept the publicity constraint shows that moral theorists, depending on their theoretical commitments, often hold radically divergent conceptions of the aim of moral theory. A key feature of motive-consequentialism, then, is its insistence on truth as the aim of a moral theory. Hence, motive-consequentialism contains a normatively neutral conception of the aim of moral theory, appealing to a simple standard of truth, with no indication that the notion of truth involved differs metaphysically from the notion of truth that might be endorsed in mathematics, or in science more broadly (see also Parfit 2011b, p. 501, where Parfit argues that moral knowledge is essentially like mathematical knowledge). The aim of moral theory is to discover moral truths that exist prior to the theory. Contractualist moral theories, in contrast, typically hold that the aim of moral theory is not normatively neutral. The mere pursuit of a moral theory presupposes that the moral agents deliberating already accept certain values. For instance, the contractualist approach to moral theory works based on the assumption that it is moral agents who are already committed to being a part of a normative agreement about the principles of morality who ask the question of what principles should govern their relations, and that publicity counts as a desideratum on moral theories because an esoteric moral theory would brush up against this foundational assumption. This is thought to constrain the principles we might adopt, as has been argued at length by Rawls and his followers (Larmore 2003, pp. 369-375).

The fact that the contractualist's position on the esoteric morality question presupposes values is marshalled into an objection to the publicity condition by Eggleston, who argues that this position requires the contractualist to simply presuppose the truth of their theory, thereby begging the most important question against the consequentialist (Eggleston 2013, p. 42). Eggleston's argument is important because it highlights the toothlessness of the contractualist view on the matter. Consequentialists like Parfit have gone to great pains to demonstrate the incoherence of the contractualist's Kantian conception of the person as precommitted to certain normative projects in virtue of their common humanity. If we look at the metaphysical facts, according to the consequentialist, there is no such thing as personal identity, and hence no such entity that can serve as the foundation for a moral theory (see Parfit 1984, p. 242 for just one example of this kind of argument). So even if one finds the contractualist's conception of the aims of moral theory to be convincing, it is not clear that the contractualist can substantiate their own case in a broader sense, if doing so requires debunking the consequentialist's arguments against personal identity (although Korsgaard (1989) has tried valiantly). For doing that would require taking exactly the kind of stand on controversial metaphysical questions that the contractualist, at least the contractualist with a Rawlsian stripe, argues that we ought to be silent on in our justification of moral principles (see also the relevant discussions in Chapters 2 and 4). From this it follows that the contractualist's criticism of the consequentialist is not a genuine one.

If the publicity constraint and the objection from esoteric morality have no teeth as an argument against consequentialism, the idea that there is a problem about esoteric morality seems to dissolve. It is not a point in favor of contractualism that it requires publicity, and it is not a point against consequentialism that it is esoteric. It is simply a feature of those views that

they take these positions. But although it is a feature of those views, it is not an unimportant one. The position that one's moral theory takes in response to the possibility of esoteric morality represents the concerns that one has when one constructs a theory. We can say plausibly that the attitude that one takes toward this reflects the concerns of one's moral theory: consequentialism is concerned with optimal design, and hence takes no concern to deontological values like publicity unless those have consequentialist relevance, and contractualism, concerned as it is with openness, deliberation, and whether a moral principle is acceptable to all parties, is antithetical to esoteric moralities.

Further supporting the claim that the problem dissolves is the simple fact that consequentialism may be derived regardless of the conception of the aim of moral theory which we endorse. Pursuing a moral theory among Parfitian lines might be a reasonable pursuit, even granting that moral theory also aims at providing the basis for public agreement, and pursuing a moral theory along contractualist lines could be reasonable, even on the assumption that moral theory sometimes aims at truth. We should also note that it's not even uncontroversial among consequentialists that it is the aim of a moral theory to specify moral truths: R. M. Hare (1981, pp. 35-39), for instance, endorses a version of consequentialism according to which moral theories aim at the reconciliation of various complex disputes, and does so without endorsing Parfit's ambitious notion of the conception of moral theory. If it turns out that consequentialism is derivable given the aim of some other conception of moral theory, then even if it turns out that contractualist views can't substantiate a conception of the aim of moral theory, then it does not follow that we should accept Parfit's conception of the aims of moral theory. But if we don't have that conception, then it's not clear how multi-level consequentialism is not subject to the esoteric morality objection, because the defense of multi-level consequentialism against this

objection presupposed that it was clearly specified that the aim of moral theory is to specify the moral truth.

To conclude this section: It seems like there is no ultimate issue about whether there is an objection from esoteric morality. This objection crucially fails, and shows that the esoteric-public dimension is not an important one in determining which moral theory we should follow. The only way to show that it succeeds is to show that moral theory is, as the contractualist thinks, solely aimed at providing principles that form the basis of a moral agreement among free and equal parties, which seems unlikely. The result of this is that we should feel free to accept esoteric moralities. What comes next is an exploration of how moral theory should respond to this fact. In doing this, I appeal to the Rortian pragmatist conception of the aims of moral vocabularies.

1.3. Rortian neopragmatism and the esoteric liberal ironist

Pragmatism suggests that esoteric moralities can live together with the demand for publicity. This is because pragmatism provides a framework for ethical reconstruction, and hence the possibility of reconciliation of competing conceptions of the aims of moral theory. Recall that I said in the introduction that pragmatism aims at explicating philosophy as a tool for solving practical problems. Extrapolating a bit from this, I want to say that pragmatism is an approach to philosophical problems that emphasizes the role of philosophical inquiry as ultimately aiming at enabling engagement with problems in the world, rather than as trying to construct philosophical systems that get at the truth in their intended domains. From this perspective, and in light of the further fact that different situations place correspondingly different demands on moral theory, pragmatism tells us to try to parse the different requirements and aims that situations demand of us, and then judge whether esoteric or transparent moral theory is most recommended. I don't

pursue a general theory of how we ought to negotiate this dynamic, because answering it would require psychological and sociological knowledge that I don't have. But I can gesture at a few things. Maybe, for example, we demand publicity when construing a social contract; this claim, of course, makes sense in light of the interpretation as Rawls's veil of ignorance as a procedure aimed at bringing discourse, as much as is possible, to publicity. And we might demand an esoteric morality when dealing with young people, or people under extreme situational demands, because an esoteric morality allows us to teach them rules of conduct designed to help them achieve their narrowly-construed goals, for children, surviving childhood and learning a more mature conception of morality, and for adults under extreme situational demands, tools to make difficult decisions that approximate, but do not correspond to, the real, better-justified moral theory. In view of this, pragmatism basically tells us to be guided by the esoteric-public dimension of moral theory when doing that would have pragmatic value. And this, in turn, helps us to see why the difference between consequentialism and contractualism may not mean quite as much as we think it does.

However, the pragmatist is not out of the woods yet. I will argue that something like the problem of esoteric moralities arises for the pragmatist, and that this problem demands an answer. (Spoiler: the pragmatist can provide at least the beginnings of one.) Before I can explain this objection, though, I want to comment on a crucial technical move that will inform what is to come.

The discussion so far has taken place in terms of competing moral theories, but the move to a pragmatist reconstruction shifts the object of discussion from theories to vocabularies. I discussed the idea of a vocabulary in the Introduction, so readers may want to remind themselves of some of those things. A vocabulary is far broader than a theory, at least potentially, and a

vocabulary might include multiple theories, along with some account of how they fit together. This shift is important because, for the pragmatist, a theory can retain its usefulness even without anything like a pretension of an appeal to truth. To restate the claims about a pragmatist balancing of esoteric and public moralities in the terminology of vocabulary: there can be a moral vocabulary, on pragmatist terms, that includes both an esoteric and a public theory, along with an account of how they fit together.

On first pass, therefore, one might think that Rorty should simply treat vocabularies as situationally-contingent, and hence think that there is no interesting philosophical issue as to whether there should be widespread esoteric moralities. This could be seen as following from his arguments that there are no final vocabularies (Ramberg 2009, sec. 2.1). Recall that a final vocabulary, were there such a thing, would be a set of meanings, theories, and conceptual presuppositions that give the final word on what there is in some aspect of reality. A non-final vocabulary, or just a vocabulary, is a set of meanings, theories, and conceptual presuppositions that provide a way to talk about some aspect of reality. For example, physics provides a vocabulary, because it brings along with it terminology, theories, and concepts that we use to make sense of the world around us. Consequentialism provides a vocabulary but in a narrower sense than physics, because consequentialism's vocabulary of consequences and the evaluation of them only brings along meanings and such with respect to morally-valenced situations. Physics could be a contender for a final vocabulary full-stop, while consequentialism could at best be a contender for final moral vocabulary. Rorty thinks that there are no final vocabularies, because all attempts to specify final vocabularies have failed, thereby raising tricky questions over what criteria a vocabulary would have to meet in order to actually be final, or because attempts have succeeded only at the price of being final only in a very narrow and uninteresting

sense (Ramberg 2009, id). Consider Rorty and Quine in conversation regarding the latter's attempt to elevate physics as the final vocabulary of all inquiry. According to Rorty, Quine gives an account of meaning whereby meaning is purely extensional, in that meanings are specified by the behavioral responses that subjects make toward queries about the meanings of terms (Quine 1960, p. 34; Rorty 1979, p. 204). As it turns out, this is a subtle misreading of Quine, although not wholly unreasonable. This view comes in the wake of Quine's arguments, in "Two Dogmas of Empiricism," that there are no intensional meanings. Intensional meanings are, roughly, meanings that are specified by the internal contents of the term whose meaning is in question. For Quine, the argument is that intensional meanings presuppose that there is a philosophically-valid account of synonymy of meanings, which is not possible except by arguing in a circle (Quine 1951, pp. 21-25). But clearly, we need some account of meanings. As it turns out, though, Quine holds that there cannot be an extensional account of meanings either, insofar as extension's being leveraged for a semantics would require the same notion of synonymy. In a later piece, "Notes on the Theory of Reference," Quine addresses this issue by arguing that, although strictly-speaking there is no philosophically-valid account of an extensional semantics that does not fall to the problem raised in "Two Dogmas," something like a semantics can be supplied by appealing to the Tarski disquotational schema, which is a "technical construction" (Quine 1961, p. 138) that allows us to define the predicate 'true-in-L', where L is some language, in a way that does not presuppose a notion of synonymy. The disquotational schema does this by essentially replacing the theory of meaning with the theory of reference, by defining 'true-in-L', for a language, as a system of mapping sentences, treated as names, onto statements in the language itself (Quine 1961, p. 137); the mapping is explained by the name and the statement sharing an extension (Quine 1961, p. 132). In a crucial moment, Quine justifies replacing

semantics with the theory of reference just because the ideas of the theory of reference, as he sees them, “are so very much less foggy and mysterious than the notions belonging to the theory of meaning,” (Quine 1961, pp. 137-8), an idea that proved influential to Donald Davidson as he further built out Quine’s notion of reference in “Truth and Meaning” (Davidson 1967, pp. 309-311). Hence, we can say that, although Quine does not think that there is an extensional account of semantics, the thing that we can do that most approaches an account of semantics is a disquotational approach to the truth predicate which itself presupposes something like co-extension. Davidson’s comments in that essay are crucial for understanding another important feature of Quine’s account: the claim that this theory is “an empirical theory,” and that “it may be tested by comparing some of its consequences with the facts” (Davidson 1967, p. 311). The question of what words mean is crucially bound up with empirical investigations into the nature of the world. As a consequence, science plays a crucial role in the disquotational approach, insofar as physics is the most general empirical investigation and provides the final word, in some sense, into the empirical investigations that can explain name-statement mappings. Physics is as close as we can get to providing a final vocabulary for all inquiry, and hence the source of an extensional account of meaning that allows us to do what we want with our language, because physics is the most sophisticated and nuanced method of gathering empirical evidence; it is the best way to be sure that our statements and truth-claims conform to observation (Quine 1960, pp. 17-21).

In response, Rorty argues that, even if Quine were right and that physics was the final vocabulary of all inquiry, in that it gave the final, maximally inclusive word on all aspects of reality, it doesn’t follow that physics is exhaustive. This argument would show that physics is universal, but it does not show that all other ways of speaking, including the vocabulary of

intensional semantics, are empty (Rorty 1979, p. 207). Even if everything that were relevant, including all ethical claims, could be redescribed in the vocabulary of physics, it doesn't follow that we have no reasons to employ other vocabularies. In other words, Rorty holds that the sense in which physics is final, even granting Quine's arguments that it is final, is not the interesting sense of finality, hinting at a distinction between a final vocabulary as one that covers all conceivable cases and a final vocabulary as simply the one that we have most recently agreed to hold constant.

The fact of the impossibility of a final vocabulary, except in the relevant less interesting sense, suggests a pragmatist attitude toward esoteric moralities. An esoteric moral theory can be a constituent part of a particular moral vocabulary precisely because the world sometimes, but does not always, demand that we employ esoteric moralities. Hence, an esoteric moral theory, like any other moral theory, would be poorly-suited as a final vocabulary, but perfectly plausible as a constituent part of some carefully worked out and continuously revised ethical vocabulary. The issue with esoteric moralities, and the issue that dominated the first section of this chapter, is simply that we have no way to say, as a matter of final vocabulary, whether the right moral theory is really esoteric, or really public, because we have no answer to the question of the aims of moral theory. Publicity and esotericness are simply constituent properties that our broader ethical vocabularies have to prepare to respond to. It is pragmatically good, we might say, if our moral vocabularies aim to achieve a complex balancing of the various aims of moral theory, rather than trying to impose a singular conception.

Note, further, that this discussion opens up the possibility of a reconstruction of rule-consequentialism along pragmatist lines. By showing that it is good, from a pragmatic perspective, to allow our moral theories to have an esoteric moral component, pragmatism

provides an attractive explanation for what appears to be a significant bullet for the consequentialist to bite. It looks like a bullet to bite because consequentialists see themselves as playing this realist game of showing that their moral theory is true, and therefore superior to deontological theories like contractualism. For instance, if consequentialism can be successfully defended against the objection from esoteric morality as Parfit thinks it can, then the aim of moral theory is to specify moral truths. Since it is not clear what the aim of moral theory might be besides this, it is plausible to say that, for Parfit, the aim of moral theory as specification of moral truth needs to be part of the final vocabulary of moral theory. It must be part of the theory's set of meanings, theories, and conceptual presuppositions that have the final word about some aspect of reality. But Rorty should be committed to denying this, and this is a good thing for the consequentialist, because it sure seems like we use moral theory with some other kind of aim at some point. Contractualism's insistence on publicity and its relationship to the social contract is an example of this. Further, even granting that Parfit's conception of the aim of moral theory could, somehow, provide a normative vocabulary for all cases, and in that sense could be a final vocabulary, that doesn't exclude the possibility that there are other vocabularies that can also be used to make sense of some case. Parfit, then, could claim to have a final vocabulary, but only in the less interesting sense in which Quine's conception of physics provides a final vocabulary. So even while being agnostic on the question of whether or not a moral theory aims, or should aim, at specifying moral truths, it can be shown that consequentialism does not provide a final vocabulary for the proper aim of moral theory, except in an uninteresting sense that it can cover all cases non-exhaustively. But appealing to pragmatism allows this problematic to drift away, and to instead see this feature of consequentialism as a feature of the fact that there are multiple aims for moral theory. The multiple levels, of this consequentialism, of providing

general principles (act-consequentialism), say, and explicating a way of life that can actualize those principles in persons and their psychologies (motive-consequentialism), are unified by Rortian pragmatism.

Whether this actually is pragmatically good, however, is an legitimate, open question. This is a significant problem for the pragmatist attitude toward esoteric moralities, because it depends on esoteric morality's being pragmatically valuable. Consequently, if there is some reason to doubt the pragmatic value of esoteric morality, then there are doubts about the pragmatist reconstruction of esoteric moralities, and hence multi-level consequentialism, that have been offered. We will explore this possibility in the next section. For the moment, however, I want to provide one more argument that pragmatism and consequentialism do a better job at dealing with these esoteric moralities. Consider the case of esoteric moralities and moral education. Adrian M. S. Piper argues that moral education demonstrates the necessity of the publicity constraint, insofar as moral education is esoteric and in virtue of that uniquely justified in the case of morality for children. We often teach children patently false moral rules, and we do this because they have intellectual constraints that would prevent them from learning the right rules. Piper argues that this shows that esoteric moral theories are problematic because we do not teach people a patently false moral rule when we consider them to be intellectually sophisticated enough to learn the true moral rule. Esoteric moralities, hence, treat grown adults as though they are children. This in turn supports the objection from esoteric morality, insofar as it supports the controversial premise of that argument that we should reject esoteric moralities (Piper 1978, pp. 205-6).

I actually think this argument doesn't support the objection from esoteric morality. In particular, it seems hard to say that we should never treat grown adults in the way that we treat

children. For instance, it might be fine to treat a person in the midst of a panic attack as a child, by cuddling them and speaking to them in simple, soothing tones. This effect doesn't seem to dissolve as choices become more consequential. If L. A. Paul's discussions of transformative experiences are to be believed, then we are often in a very poor epistemic situation with regard to major life decisions that we make, insofar as these decisions affect who we are, and hence deprive us of the subjective experiences that would explain whether that change was right (Paul 2014, esp. pp. 13-14). In those cases, some measure of self-deception might be justified, or even unavoidable. It seems to follow that a person undertaking a transformative experience might adopt a moral theory that allows self-deception. This is all to say that Piper's conception of publicity is overly strong. It is simply not true that we should always be transparent with grown adults, and correspondingly it doesn't seem always objectionable to treat grown adults like children. The fact that in some cases it would be wrong to treat adults like this is also irrelevant, because a pragmatist can simply say that theory aims at different things in different contexts. With some more nuance as to what it means to say that we should never treat grown adults as children, we see that adults and children are relevantly similar, in at least a narrowly-circumscribed set of cases, and hence that we can't support the objection from esoteric morality by arguing that we should always reject esoteric moralities because they treat grown adults as children.

Pragmatism and consequentialism explain these conclusions better than a moral theory that meets the publicity constraint. A moral theory might aim at a pattern of being believed that corresponds to the level of intellectual capacity for differing individuals, insofar as intellectual capacity affects the aim to which a person would put the moral theory. We simply allow that moral theory aims at what a person, at whatever level of intellectual competence they are at,

needs from a moral theory. Hence, contra Piper, there is no reason to think that moral education supports the esoteric morality objection. Actually, moral education supports its pragmatist dissolution.

1.4. The esoteric character of conceptual feminism

Now we know what it is to dissolve the esoteric morality objection by Rortian pragmatism, and we have a test case, drawn from the literature on esoteric morality, which shows how this could be a good thing for moral theory. Yet with all that has been said, Rorty is not out of the woods on esoteric morality. Rorty thinks that, when considering our ethical vocabularies, we have to adopt the standpoint of the liberal ironist, a citizen of a democratic society who remains committed to their own final vocabulary, and yet realizes that it is a merely contingent fact that they have the final vocabulary they do. This person “combine[s] commitment with a sense of the contingency of their own commitment” (Rorty 1989, p. 61). A liberal ironist has the uncomfortable project of trying to balance maximal openness to reconsideration of their vocabulary while still remaining committed to that vocabulary. This fact gives rise to a possible objection to Rorty’s response to EM. Requiring theorists who tend to be underrepresented in moral theorizing to be liberal ironists could undercut a different legitimate aim of moral theory: providing a toolkit for underrepresented individuals to analyze and critique their experiences of injustice. If Rorty’s position on esoteric morality ultimately makes it harder to assert the finality of one’s social justice vocabulary, and being able to assert this has pragmatic value, then this would show that Rorty’s position is bad, from the perspective of pragmatic value. In this section, I lay out this objection in more detail, and show how Rorty’s ideas give us the resources to account for it. In doing that, I show how a Rortian pragmatist dissolution of the objection from

esoteric morality helps us think clearly about the aims of moral theory in high-stakes real world cases.

Let's grant, for the sake of argument, that a Rortian pragmatism allows for multiple competing conceptions of the aim of moral theory. If my claims about Rorty's pragmatism in the previous paragraph are correct, then it seems like that is just a simple reading off of his view. This, however, would entail a substantive commitment on the part of Rorty for moral agents to adhere to a broad variety of moral vocabularies, with this adherence determined by the situational and dialogical on agents. If this is true, then an agent might face an esoteric morality insofar as situations may make differing, incompatible demands of agents.

Suppose one thinks that we ought to use conceptual ethics, conceptual engineering, and metalinguistic negotiation to develop a sharper, more practically-robust vocabulary for social justice projects. Conceptual ethics is the field of ethics that investigates, to quote Herman Cappelen and David Plunkett from a recent review article, "normative issues about which concepts one should use (and why) and evaluative issues about which concepts are better than others (and why)" (Cappelen and Plunkett 2020, p. 4). In this article, I focus on the work of Sally Haslanger and numerous philosophers, who have held that certain social justice concepts like race and gender (Haslanger 2005, pp. 20-21), misogyny (Manne 2016, p. 44), disability (Barnes 2016, pp. 39-48), and health (Kukla 2015, pp. 525ff)¹ are defined in common parlance in a way that robs them of the critical import for which they are needed. A good example is Kate Manne's argument that the term misogyny is defined in common parlance as a universal psychological disposition to hate women and to act so as to harm them, that this definition forecloses the use of the term for the crucial social justice purposes for which feminists need it, and that we therefore

¹ It should be noted that, although it's quite clear that the approach in this paper mirrors that of Haslanger and the others, Kukla does not explicitly use the language proposed by these authors.

have reason to push for its redefinition to be more sensitive to the social and political components of misogyny (Manne 2016, p. 44). This feminist push for the redefinition of social justice terms for specifically feminist reasons I will call conceptual feminism.

It is not obvious how conceptual feminism relates to moral theory. After all, moral theories usually tell us what we should do, not how we should talk. One way to understand how this could be is to consider moral theory as an attempt to achieve reflective equilibrium between our considered moral judgments, moral principles, and, in a crucial addition made by Norman Daniels, our broader background theories, including “theories of the person or of the role of morality” that govern the predictions that we make about our normative world (Daniels 1979, p. 261). If we count theories about language and semantics as a part of our broader social and psychological theories, then it seems like giving a complete moral theory would need to answer to the possibility of conceptual engineering. The social and psychological theories might, for instance, have to contain an account of the rules, both descriptive and normative, of the use of the method of conceptual engineering. Daniels, broadly concurring with this, notes that our background theories might lead us to reject a principle or a considered judgment, but that a principle or a considered judgment might also, possibly counterintuitively, lead us to reject a background theory (Daniels 1979, p. 259). It hence seems that conceptual feminism allows us to occupy a sort of second-order control over moral theorizing, insofar as it specifies possible conditions under which intentional modifications to our language can affect the background rules that constrain normative theorizing, and which normative theorizing must be accountable to.

The possibility of conceptual ethics and a Haslanger-style redefinition of social justice concepts raises a problem for a Rortian attempt to dissolve the concern over esoteric moralities. Suppose we think that moral theory’s aim is to pursue new vocabularies, to provide vocabularies

that are inclusive of a maximal number of perspectives, and to destabilize vocabularies that are proposed as final vocabularies. A conceptual engineering project on the concept of misogyny is in effect a modification of our moral vocabulary. Which aim of moral theory does this satisfy? It doesn't take much effort to see that it satisfies some, and fails others. Conceived as the aim of pursuing new vocabularies and including the maximal number of perspectives, Rorty would surely be welcoming of conceptual projects like Manne's. Yet if moral theory aims to destabilize putative final vocabularies, Rorty has to oppose Manne's project, insofar as Manne is giving an analysis of feminist concepts that allows feminists to exclude various other analyses of those concepts for social and political reasons. As Nancy Fraser (1991, p. 265) writes, the issue with Rorty on these points is that he both requires an account of who has this semantic authority, given his explicit commitment to social justice projects, and provides us with an approach to moral philosophy that cannot make sense of the question itself. Presumably the feminists on his view ought to hold this semantics authority, but Rorty's esoteric approach to morality seemingly requires that this role can only be taken either illicitly, or by someone in a private sphere, cut off from a broader discursive community that must be targeted by the project.

The possibility of conceptual ethics shows, I think, that the broad array of pragmatist conceptions of the aims of moral theory pull in different directions in certain contexts. This problem is mirrored in some arguments given by Margaret Urban Walker (1996, pp. 284-287), who argues that feminist philosophers should oppose esoteric moral theories because of the tendency for an esoteric morality to encourage domination of women because it hides crucial features of their own experiences of gender oppression from them. To provide women with esoteric moralities is to position them in such a way that makes them less likely to participate in moral discourse, and hence more likely that we end up with moralities that reinforce patriarchy.

Rorty, it could be thought, winds up with exactly the kind of problematic morality that Walker repudiates. In an essay called “Feminism and Pragmatism”, Rorty holds that the aim of destabilizing patriarchal final vocabularies is among the crucial aims of feminist theory (Rorty 1998b, p. 8). Conceptual engineering projects, however, show that there is a social justice reason to provide final vocabularies whose destabilization we actively oppose. Moreover, the reason that we should actively oppose this destabilization is provided by an uncontroversial conception of pragmatism. So the pragmatist is tied up in knots, on the one hand with a theory that tells them to work toward justice and equity for women, and on the other with a metaphilosophical outlook that requires openness to vocabulary modification that could undermine that aim, insofar as it allows anyone, up to James’s ‘last man’, to participate.

The dominant theme of Rorty’s essay is that pragmatist opposition to universalism and realism couple well with, and gives a satisfying account of, the feminist critique of commonsensical social structures as patriarchal. In writing this, Rorty rightly captures crucial facts about these sorts of attempts. Yet when the moment comes to consider what it might take to implement feminist discourse, Rorty writes the following: “Although practical politics will doubtless often require feminists to speak with the universalist vulgar, they might profit from thinking with the pragmatists” (Rorty 1998b, p. 210). This claim is astonishing for a pragmatist who talks like Rorty, for it in effect proposes the creation of a set of norms and practices for universalist discourse as an aim of feminist theory. If, as seems fair to say, this in effect proposes a new pragmatist notion of the aim of moral theory, then it seems that Rorty has endorsed aims of moral theory that pull in opposite directions. If Rorty is serious about his allegiance to feminism, then this problem demands an answer.

I claim that this example shows that esoteric morality seeps in through the back door, even for a pragmatist who attempts to defuse theoretical dilemmas like the esoteric morality problem. For even if a pragmatist has the aims of destabilizing vocabularies, the aim of destabilizing vocabularies may tell a person to come to have beliefs whereby certain vocabularies are stabilized. If Rorty is right that practical politics requires a stable vocabulary of social justice concepts, as he seems to be forced to by his claim about the ‘universalist vulgar’, then insisting on the destabilization of all vocabularies, which in effect treats all attempts to stabilize vocabularies as attempts to turn a vocabulary into a final vocabulary, could undercut feminist theory exactly where it needs to be elevated.

A similar kind of argument is made against Rorty by Bernard Williams. In *Truth and Truthfulness: An Essay in Genealogy* (2002), Williams argues that there are things called everyday truths, certain kinds of clear truth-claims that we are willing to accept without much critical analysis. The everyday truths include truths about the weather and about what things we can plainly see; indeed, Williams also refers to these truths as plain truths. One way to read Manne’s account of misogyny, and indeed conceptual engineering claims in general, is as an attempt to influence what should count as everyday truths about misogyny. For example, Manne may be read as advocating that we should understand misogyny in such a way as to say that sentences like ‘Misogyny need not be intentional’ are everyday truths (c.f. Manne 2016, p. 61). Now Rorty, given his explicit endorsement of the feminist project, should at least in theory be amenable to the idea that feminists ought to exert control in this way; Rorty, however, denies feminists like Manne the ability to assert everyday truths about misogyny, because as per Williams his approach to truth specifically, and moral theory more generally, is thoroughly ironist, and hence indifferent to whether there is such a thing as everyday truth. Paraphrasing

Williams, Rorty's pragmatist approach to truth undermines the position that the critic attempts to take, because it both requires that certain everyday truths about misogyny be asserted, and that there are no truths, save for the 'truths' of power (Williams 2002, pp. 8-9; see also Fricker 2013, pp. 811-12; Lovibond 2015b, p. 48).²

Again, taking Manne's discussions of misogyny as an example, we could read the import of that project at the attempt to impose a normative vocabulary, which specifically acts as a tool to hammer out socially- and culturally-mediated disputes about the semantics of the term. It is, of course, a matter of social-political strategy that Manne appeals to here. She writes that her conceptual ethics project is necessary, given how the term misogyny is currently used, to avoid silencing women in the attempts to articulate the conditions of oppression (Manne 2016, p. 44). On such a view, the utility of the social-political conception of misogyny is in part found in the fact that it aims to exclude the individualistic vocabulary from common parlance. What Manne does not argue, however, is that appealing to an individualistic account of misogyny is not the only way to silence. Any assertion that undermines what we might call the practical force of the social-political conception of misogyny might count as silencing, insofar as Manne's conception, it is argued, is necessary for feminist social justice projects. Yet it could plausibly be argued that a Rortian pragmatism, targeted at destabilizing final vocabularies, undermines that practical force (Lovibond 2015b, Fricker 2013). This follows because Manne's project in effect aims at arguing for the primacy of some vocabulary. In arguing that we have reasons to accept the social-political conception of misogyny over the individualistic conception, Manne is committed to, if not the claim that her preferred conception is the final vocabulary of the matter, the claim that there is an objective ordering of vocabularies over which she has semantic authority, which

² I am indebted to Miranda Fricker for calling my attention to these texts, and for putting this objection to me.

answers to a normative conception outside of those vocabularies. If arguing for the finality of a social justice vocabulary is what Manne aims at, then Manne's aim runs afoul of the broader conception of the aims of moral theory that Rorty endorses. But given Rorty's support for feminist projects, this result is unwelcome.

For Rorty, however, this problem is not insurmountable, and seeing how we address it provides a helpful example of the possibility of esoteric discourse which can be seen to be acceptable to both those engaged in feminist social justice projects and a Rortian who wants to provide a robust account of how moral theory, or something like it, can be philosophically responsible, on pragmatist lines. If moral theory serves multiple, divergent aims, each themselves a part of a differing moral vocabulary, then a suitable theory of social justice concepts will require discourse of a different tenor depending on the demands of the situation, and depending on the abilities of the people deliberating within the situation. Feminists of Manne's stripe need an objective ordering of vocabularies, because this objective ordering is itself necessary to retain the assertoric force of social justice terms. But the broader aims of feminism as a critical philosophy must hold the non-absoluteness of whatever objective ordering is endorsed. It is hence a mistake to assume that the fact that there are multiple aims of moral theory makes any trouble for conceptual feminism as it exists in the case of the pragmatist reconstruction of motive-consequentialism. It is no more troubling than the fact that different audiences, and different situations, require theorists to engage in discourse that aims at different things; the multiple aims are themselves a consequence of the non-absoluteness of the feminist vocabulary. Similarly, with the Williams point, it is simply not right to say that pragmatism like Rorty's can offer us no everyday truths about which feminist claims need to be elevated, and it is simply not right to argue that this is because pragmatists are too relaxed toward truth-claims. Pragmatists

rather recognize that discourse and persuasion are deeply contextual, and that any attempt to state everyday truths about how to pursue social justice projects needs to be placed within the context in which the claim is to be made. To say that these contextualized everyday truths are genuine truth claims, therefore, can be asserted by pragmatists precisely because it is vacuous, and something that anybody can get on the cheap. But it is not clear what more would be required in order to drive a social justice project; there is a crucial difference between a truth-claim and a claim about what reasons people have, and the latter is what really matters. This might also be true at the same time as claims about the reasons that people have can be said to be true. For instance, if someone is punching me, I can tell them to stop punching me, and whether or not they recognize that reason, saying that it is true that the person should stop punching me adds nothing to the reason that the person has to recognize the reason. It is an everyday truth, I think, that the person should stop punching me, but adding the truth-claim to this adds nothing to make that everyday truth more compelling, and if we can find a value in adding the truth-claim, so much the better for the pragmatist, as the value would be pragmatic value!³ (See Chapter 4 for a more detailed discussion of the pragmatist approach to truth, and why these claims can be made.) Feminist theorists, at least potentially, need a vocabulary for engaging women's problems that both allows feminist criticism of real practices and for the vocabulary in which that criticism is couched to be the possible aim of reconstruction, whereas women going about their lives outside of theory need a vocabulary that places that line differently, and privileges an approach that prioritizes the enabling of feminist criticism of real practices. Which of these vocabularies is more true or not is not to the point; what matters is whether the selected vocabulary is more effective at getting the punching to stop. Note, then, that showing that a pragmatist feminism

³ I am indebted to Jesse Prinz for examples of this kind.

endorses an esoteric, ironist conception of the varying levels of feminist discourse is no objection to that version of pragmatist feminism. Rather than dividing feminists against themselves, a suitably interpreted conception of Rortian pragmatism explains and supports the subtle dialectical moves that feminists make to further their causes.

1.5. Weaving together the threads

The puzzle about what we should say in the face of moral theories that license esoteric beliefs hence cannot be resolved by taking a pragmatist stance toward the possibility of moral theory, but this is not a problem. The puzzle arises because moral theories purportedly aim at truth, and truth and belief are not necessarily linked. But even giving up the idea that moral theories aim at truth can give rise to a puzzle about esoteric moralities, because a moral theory might have multiple, conflicting aims. Of course, although the problem arises for pragmatists as though it does realists, the problem does not come with the same kind of stigma that it does for a realist. There is no reason to think that there should be neat answers to these problems, and even the idea of proposing a moral theory in the first place is hard to square with pragmatist framing assumptions. Yet even on all these claims, a theoretical approach to morality along pragmatist lines might tell people to disbelieve in pragmatism, for the purpose of practical politics. This is a kind of problem of esoteric morality, although again, it is not as serious a problem for the pragmatist. There is no reason to think that a pragmatist theory needs to be true or believed in, and so it's not surprising that practically, we can negotiate the strength of that belief when it suits our purposes, as when conceptual feminists aggressively insist on an engineered analysis of social justice terms. But it does show that even the pragmatist's attempt to avoid this knotty philosophical problem cannot succeed so simply as they'd hope.

These arguments show that pragmatism, in the character of the liberal ironist, conceptual feminism, and multi-level consequentialism share an internal dialectical structure. Each is an attempt to negotiate and substantiate the various aims and needs which motivate moral agents to utilize moral theories across time and variable contexts. The idea that we can reject a multi-level consequentialism by appealing to a publicity criterion, at the end of this, appears even less plausible than it did at first, because appealing to a publicity criterion would make the position of the liberal ironist and the conceptual feminist even harder to maintain. But the world demands of us that we maintain these positions, insofar as we are sophisticated creatures that aim at radically different things depending on our contexts. Hence, we have a pragmatist reason, albeit an indirect one, to assent to the standpoint of esoteric, multi-level consequentialism, if that is the path we choose to take.

This does not amount to a real, direct defense of consequentialism, however. If it were direct, then the contractualist would not be able to reconstruct their own position along pragmatist lines. But, as I will argue in the next chapter, contractualists, by appealing to the broader possibilities that are available in the literature on deontology, can reconstruct their theories in much the same way, by appealing to pragmatist framing conceptions about how the practices of moral theory are best internally structured and engaged in. The notion of internal structure is there leveraged to provide a multi-level deontology that reconciles various considerations proposed by deontologists whose projects, on their faces, are radically divergent in aims and emphasis.

CHAPTER 2: Reconstructing Moderate Deontology in Political Liberalism

2.1. Introduction: moderate and threshold deontologies

The previous chapter tried to show that pragmatism can vindicate an important aspect of the consequentialist perspective, and hence that pragmatism can raise problems for deontological theories like contractualism. Yet this does not amount to a refutation of deontology, because the goal in raising those problems was simply to show that a plausible pragmatist reconstruction of consequentialism is possible. Consequentialism is a position that a liberal ironist can endorse, use, and defend. It is the project of this chapter to show that something similar can be done for deontology, in view of the many problems involved in formulating the multi-level deontological theory known as moderate deontology.

It is commonly held that deontologists think that it is impermissible for a physician to kill a healthy patient and part out the patient's body parts for transplant into five other patients, whose lives would be saved. Call this the 5-to-1 Case. This is because deontology is a system of ethics that typically holds that it is always impermissible to do certain actions that disrespect the fundamental moral equality of persons. Killing that one healthy patient seems to do just that. In deontological jargon, there are *deontological constraints* against performing certain actions, and these constraints always hold, with no exceptions. Importantly, deontology apparently does not allow consequence-based exceptions to constraints. Deontological constraints are supposed to hold regardless of the consequences of respecting them (Alexander and Moore 2020, sec. 2).

Most deontologists, even those who endorse an absolutist conception of deontological constraints, recognize that extreme cases put pressure on this standard conception of constraints. In a case where a physician could kill one healthy patient and part out that patient's body parts for transplant into 5,000,000 other patients, whose lives would be saved, it seems that the

impetus to permit that action is stronger than in the 5-to-1 Case. Call this variation the 5,000,000-to-1 Case. I assume, for the sake of argument, that the reader has the intuition that killing in the 5-to-1 Case would be impermissible, and that killing in the 5,000,000-to-1 Case would be permissible.

This pattern of judgments raises the problem of *moderate deontology* (Smilansky 2003), alternatively called threshold deontology (Alexander 2000, pp. 893-4; Brennan 2009, p. 26), non-absolutist deontology (Ellis 1992, pp. 857-8), and weak deontology (Huemer 2009, pp. 464-5). If killing is impermissible in 5-to-1 but permissible in 5,000,000-to-1, as intuition would have it, then there must be some number at which killing switches from being impermissible to permissible, some n such that, in n -to-1 cases it is permissible to kill, but in $(n-1)$ -to-1 cases it is impermissible to kill. In the jargon of moderate deontology, n is the *threshold* at which deontological constraints are relaxed.

The problem of moderate deontology is that the possibility of a deontological threshold makes trouble for the conception of deontological constraints proposed in the first paragraph. Deontological constraints are either *absolute* or *moderate*. If deontological constraints are absolute, as a straightforward understanding of deontology would suggest, then the intuition that killing is permissible in 5,000,000-to-1 would have to be rejected. However, if deontological constraints are moderate, and subject to thresholds at which they no longer constrain, then the intuition about 5,000,000-to-1 can be accepted. Most, as I have already said, regard this latter course as correct. But this move creates a number of significant philosophical problems. A first problem regards the question of the placement of n , the threshold. What is the value of n ? And without a conclusive answer to that question, how can we decide on a value for n ? A second problem regards the possible manipulability of deontological thresholds. If we are in a situation

where a group is one person below tipping a deontological threshold, assuming that we have placed n , is it permissible to kidnap a final person and add them to the group, thereby tipping the threshold? Intuitively this violates deontological constraints, but it is hard to see how it might be ruled out by a moderate deontologist.

Although both of these problems are interesting and will figure into this chapter at various points, a third problem is the primary topic of this chapter. Is moderate deontology genuinely deontology? Some, including Saul Smilansky, have raised this question by noting that moderate deontology is too weak on the nature of deontological constraints to count as genuinely deontological. On such a view, all someone needs to be a deontologist is to accept that there are “only some constraints” on our actions, and not absolute constraints that forbid actions that violate moral equality. Smilansky grants that moderate deontology trades on a plausible intuition for deontologists: that even granting the absoluteness of deontological constraints, people’s well-being matters quite a lot. He argues, however, that ‘moderate deontology’ is actually not deontology at all, but instead is a pluralist ethical view that gives us deontology up to a point and consequentialism after that point. But this is too permissive, says Smilansky: “We should retain our understanding of the deontological element as pure and absolute, as far as it goes” (Smilansky 2003, p. 72). Hence, while moderate deontology might be intuitively satisfying, those who see themselves as offering a defense of deontology by appealing to the existence of deontological thresholds to answer the problems of absolute deontology cannot make such a move. More specifically, Smilansky reconstructs what he takes to be the normative core of a true deontology as follows; noting that even granted deontology’s focus on the priority of deontological constraints, true deontology does say something about consequences:

[Deontology has] two sorts of valid concerns:

- (a) Deontological concerns proper (e.g. the innocent cannot deserve to be punished, hence must not be)[, and]
- (b) Consequentialist concerns (e.g. increasing human well-being would be good).

Nevertheless, and this is the determining point, the deontologist of the sort we are considering also has an underlying third principle:

- (c) (a) and (b) are lexically ordered: Whenever (a) and (b) conflict, follow (a) (e.g. if increasing human well-being in a given case depends upon punishing the innocent, this must not be done) (Smilansky 2003, p. 73).

The point is that deontology itself seems to be essentially committed to (c), and that moderate deontology is not really deontology if it denies (c), as it seems like it must. Hence, according to Smilansky, moderate deontology denies what is essential to deontology.

In this chapter, I argue that moderate deontologists, contra Smilansky, can deny (c) while still remaining deontologists. Although moderate deontology allows cases where consequences are determinative of permissible conduct, a moral theory that frames the relationship between (a) and (b) in a way that *favors*, in some technically-defined sense, (a) is still genuinely deontological. A deontological constraint can fail to be absolute, yet still rigorously constraining, and a deontological constraint can still be deontological even if not lexically prior, in Smilansky's sense, to consequences. A related conclusion that will follow is that moderate deontology is best interpreted as not being solely concerned with the problems outlined with deontological thresholds above. Moderate deontology is best interpreted along pragmatist lines, as being about designing institutions and methods for the placement of deontological thresholds that are public and that work toward moral consensus, as can be achieved within a liberal political order.

2.2. The Legal Model and Rawls's four-stage sequence

If we choose to reject (c), we need a non-absolutist conception of deontological constraints that captures enough of the nature of a deontological constraint that it is still recognizably deontological. I suggested in the previous paragraph that the way this works is that, even while rejecting (c), we can hold that our moral theory is framed in a way that favors (a) over (b), while stopping short of the overly-strong (c). It is now time to explain how this is possible, because I imagine that a proponent of (c) would simply insist that the idea of favoring simply shifts the question. What we want is an account of when it is permissible to favor (a) over (b), excepting the possibility of reverting to (c), which claims that (a) is lexically, absolutely favored over (b). If not cashed out in terms of lexical priority, then what does it even mean to favor (a) over (b)?

In this section, I show how this account is possible. To begin, consider Anthony Ellis's helpful yet under-emphasized discussion of what he calls the Legal Model of specifically threshold deontology, as an attempt to answer the placement problem (Ellis 1992, pp. 870ff). According to the Legal Model, the problem of the placement of n is addressed through a quasi-legalistic model of reasoning, whereby decisions about the placement of n are to be governed by judgments about what is an acceptable value of n , given that the aim of placing n is a part of a theory of regulating moral conduct. Any value of n is acceptable so long as it does not undermine the aim of trying to specify deontological thresholds. In this way, Ellis argues that thresholds need not be placed arbitrarily, because meeting the constraint of not undermining the aim of specifying thresholds meets the normative substance of the aim of a moral theory like threshold deontology.

It is not clear exactly why this model is called the Legal Model. The argument hinges on a conception of the aim of providing a theory of deontological thresholds, but many kinds of normative theories hinge in some important way on the aim of providing a normative theory. We might appeal to the aims of a work of art, for example, in attempting to judge some particular work. Conversely, actual legal reasoning may diverge from this kind of aim-based consideration. Certain kinds of law, such as statutes governing criminal sentencing, may be based on intuitions about the desert of criminals rather than the function of regulating behavior. It seems more germane, I propose, to call the model that Ellis proposes the Prudential Model or the Practical Model, or even the Political Model. The idea of letting the aim of moral theory determine the placement of n in threshold deontology in essence just appeals to what we would want a plausible moderate deontology to do. We want it to be able to place broadly acceptable n 's, and to do so in a way that can leave us with some confidence that we. This reading of the passage brings out a virtue of Ellis's account: that the Legal Model provides a procedure, or method, for placing n , rather than straightforwardly placing n . That is, it tells us how to figure out what n is, rather than telling us what n is. In proposing this account, Ellis frames the problem of threshold deontology as an interpretive problem. The setting of a deontological threshold is an act of approximating, in action, a judgment about what suits the purposes of having moderate deontological constraints.

As it turns out, Ellis's account survives Smilansky's challenge in virtue of this fact. Because the appeal to the aims of setting deontological thresholds provides a non-arbitrary sense in which (a) can favor (b) even if we reject (c), the threshold deontologist can admit the intuition in the 5,000,000-to-1 case and remain deontological.

In conceiving of the problem of the placement of values of n as a practical, rather than theoretical, problem, Ellis shows that we can justify rejecting (c), and hence Smilansky's concern regarding the nature of deontology, by appealing to the fact that a general rule simply stating that we occasionally choose (b) over (a) in extreme cases should not be taken to require that we give a precise account of in which cases this is justified. It is one thing to have a general principle to make sense of exceptions to deontological constraints, but an entirely different thing to design institutions and interpersonal norms that come closest to approximating that general principle. Yet in admitting that applying the general principles to the world requires those holding the principle to make exceptions to that principle does not mean that those individuals no longer accept that principle. Similarly, an institutionalized principle that governs exceptions to deontological constraints need not mean that those who deliberate about those exceptions, and sometimes act consistently with those exceptions, reject those constraints, and thereby stop being deontologists.

Another issue with Ellis's Legal Model is that it isn't nuanced enough; although Ellis's project is admirable in attempting to explain how we could bypass concerns about the precise placement of deontological thresholds, it should also matter what kinds of things would need to be done in actual societies in order to implement the Legal Model, given that placing deontological constraints will be a difficult applied philosophical task for individuals to undertake. In explaining how application works here, I suggest that we appeal to some comments by John Rawls. In an often-overlooked passage in *A Theory of Justice*, Rawls develops what he calls the four-stage sequence of principle application to govern the implementation of his theory of justice as fairness in a society (Rawls 1999, pp. 171ff). Principle application is a problem for Rawls because he thinks that the selection of the principles of justice occurs behind a veil of

ignorance, which denies to the parties to the original position particular knowledge about themselves. Depriving choosers of this information is well-motivated because it prevents them from making use of this information to design principles that could give certain majority groups advantages (Rawls 1999, p. 11). Yet it is also true that applying justice in a society will require us to know certain facts about that society, and what it needs from its institutions. So in order to apply Rawls's theory, we need to figure out a way around the informational restrictions imposed by the veil of ignorance. In the section at hand, Rawls specifies a method for gradually lifting the veil, whereby its restrictions are incrementally lifted as the level of principle application gets more and more specific.

Rawls holds that the original position, with the full veil of ignorance, is the first stage of the four-stage sequence. At this stage, parties only know the contents of what Rawls calls the circumstances of justice: general facts about social and psychological theory (Rawls 1999, p. 173). Importantly, they do not know facts such as their own institutional affiliations and the general facts about their societies. Here, the principles of justice agreed upon are not sufficiently detailed to regulate a society. Agents would need to know some details about the character of the society in question, so that the application of the principles can be tailored to what Rawls calls the "general facts about [the] society, that is, its natural circumstances and resources, its level of economic advancement and political culture, and so on" (Rawls 1999, p. 173). Rawls calls this second stage of the sequence the constitutional stage, insofar as this level of principle application roughly corresponds to the sort of reasoning that occurs at a constitutional convention. Rawls then notes that a constitution is also not sufficient to apply the principles of justice, insofar as constitutions are used to guide and restrict the "procedural arrangements" that result (Rawls 1999, p. 173). A constitution does not, in other words, determine statutes for regulating conduct.

Hence, the third stage, which Rawls calls the legislative stage, adds to the constitutional stage the task of selecting between various statutory schemes for regulating a society. Rawls mentions no changes to the position of the veil of ignorance in this case. Even laws, however, must be applied. It is a separate question, after a statutory scheme is implemented, how citizens should relate to that scheme. Hence, Rawls introduces the fourth stage, which he does not name but which has been called the administrative application stage by Colin MacLeod (2014, pp. 167-8). At this stage, the veil of ignorance is fully lifted, and all persons know all relevant facts about themselves, in addition to facts about the circumstances of their society and the circumstances of justice. Here, application is likened to “the application of rules to particular cases by judges and administrators, and the following of rules by citizens generally” (Rawls 1999, p. 175).

Citizens at the fourth stage may disagree with each other about any number of normative claims which are made, and may themselves be unsure as to which decisions would best institute the principles of justice. Hence, it seems that the four-stage sequence should lead people to accept a lack of precision in their implementations of the principles. Rawls (1999, p. 173) suggests this fact when he claims that the constitutional stage gives us the task of selecting from the many possible just and feasible arrangements the one that is most likely to give rise to just institutions. It is also not clear that demanding more precision would be the best way to institute the principles of justice. Yet citizens need not, in spite of this, lose their commitment to the principles of justice if they think this way. This is because the principles of justice are sufficiently general that one can interpret them in a number of ways and yet remain committed to them. To borrow MacLeod’s (2014, p. 169) language, so long as one is not a pluralist about the principles of justice themselves, one can be a fairly radical pluralist about the admissible possible application schemes for the principles of justice.

A further note about the four-stage sequence. If we group the four stages by the epistemic constraints placed on them, then it follows that we actually end up with three groupings: the first stage, with the full veil, the second and third stages, with a partially-raised veil, and the fourth stage, with a fully-raised veil. Supposing we were to imagine ourselves, contra Rawls's conception, as being interested in principles at all of these stages of application, it follows that there is in the procedure a tripartite taxonomy of principles, corresponding to the level of application and the epistemic specificity it allows. At the first level, we are concerned with principles at the highest level of generality, a fact forced upon us by the corresponding strength of the veil. At the second and third levels, we are concerned with principles at a middle level of generality, answering to general facts about the society in question. Then, we can consider principles at a lower level of generality, tailoring them to more specific social conditions, since we now don't have a veil of ignorance to preclude us from doing this. These principles are individual level because they are specific, and tailored to persons or groups in view of their individual needs. As we'll see, the idea of a mid-level principle is crucial to our making sense of how it could be possible to accept a moderate deontology which solves the placement problem and retains genuine commitment to deontology.

The four-stage sequence appeals to the practical infirmities of applying principles in the real world to explain how the principles can be best actualized under conditions that prevent the principle from being applied more directly. Hence, threshold deontology can appeal to a similar notion of interpretation to explain how threshold deontology can retain what is distinctive about deontology, in that it grants a significant favoritism to deontological constraints, without providing a precise formulation of the thresholds of deontological constraints. A general principle that directs us to favor (a) over (b), but stops short of accepting (c), is hence easy to

explain on a threshold deontology that adopts the Rawlsian four-stage sequence of principle interpretation.

2.3. Mid-level principles

I have argued that Rawlsian methodology can help explain how moderate deontologists can reject absolutism yet remain genuinely deontological in their moral theory. In this section, I want to further flesh out this view, by showing that this methodology can account for a crucial technical issue that has dogged threshold deontologists: the possibility of the manipulation of thresholds. Before I explain how I can solve this problem, let me rehearse the problem itself, as well as one prominent answer that has been given to it.

Suppose that we have a case before us where $n=1,000$; that is, a case where the threshold for justifiably overriding a deontological constraint against killing is 1,000 lives. We have before us, however, a group of 999 individuals. If one more person were added to the group, then the threshold would be triggered, and it would be permissible to override the constraint. Now suppose that I can kidnap a person and add them to the group. I now have before me a group of 1,000 individuals. Suppose further that I can kill one person outside that group to save all 1,000 members. Call this the 999-to-1-Plus-Kidnap Case.

Larry Alexander, who formulated the problem, argues that threshold deontologists have no principled way to say that it is impermissible to kidnap the one in this case. Yet intuitively, it seems like it is impermissible to kidnap the one. Moderate deontologists who focus on thresholds are faced with the problem of how to get the intuition to match the theory in this case (Alexander 2000, p. 903).

In response to Alexander's challenge, S. Matthew Liao has argued that threshold deontologists can hold that other deontological principles might conflict with, and override,

deontological thresholds in these cases (Liao m. s., pp. 30-1). Liao says that, in the 999-to-1-Plus-Kidnap Case, kidnapping would violate the kidnapped person's rights. Even granting that the tipping of a deontological threshold can justify harming a person as a means to an end, it does not mean that all other deontological considerations disappear. Hence, threshold deontologists can solve the problem of manipulation by accepting the intuition that it is impermissible to kidnap the one in the 999-to-1-Plus-Kidnap Case, while accounting in a principled way for why this is.

Liao further supports this conclusion by showing a case in which threshold manipulation seems to be permissible. In this case, $n=1,000$ and we have before us a group of 1,000 people who can be saved by killing one. Just as we have made up our mind and decided to kill the one to save the 1,000, one of the group escapes. Now, we have only a group of 999 people who will be saved by killing the one. The escape triggers the deontological threshold in the opposite direction; the case now falls below the threshold. Call this the 1,000-to-1-Plus Escape Case.

Liao suggests that threshold deontologists can accommodate this case. Because we are justified in killing in order to save 1,000 lives, to go through with the action in 1,000-to-1-Plus Escape would still result in the requisite number of lives having been saved. It's just that, because of contingent circumstances, one of those lives was saved in a different manner. And, crucially, it doesn't seem as if any additional rights violations occur in this case. Hence, as seems intuitive, threshold deontologists can say that it would be permissible to act in this case, and can do so for a reason that comports with the theory in a broadly satisfying way.

I don't wish to controvert Liao's solutions. The relevance of these solutions is to point out that threshold deontologists can appeal to mid-level principles in their interpretation of the general statement of threshold deontology, as requiring us to significantly favor deontological

considerations over consequentialist considerations. Corresponding to the Rawlsian four-stage sequence, these principles would fall into the level of the second and third stages of theory interpretation. Starting from the general idea of threshold deontology as a part of the first level, we then specify restrictions on the application of threshold deontology. We should significantly favor deontological constraints over consequences. Now, more specifically, we want to know how to favor deontological constraints over consequences; we want a more robust interpretive principle that tells us that further rights violations than just the one required to act according to the deontological threshold are impermissible, to help us cash this idea out. In other words, we want a mid-level interpretation of threshold deontology that captures the importance of other deontological values, such as rights. Maybe, then, something like a principle arises from Liao's response to Alexander: do not favor consequences over deontological constraints in such a way that violates rights in addition to the rights-violation that would have to be violated in order to abide by the deontological threshold. Once we know how to answer concerns about the manipulation problem in this way, we can then proceed to a further stage down the sequence, where we get clear on the problems with having actual people attempt to reason about deontological thresholds.

Once we have a grasp on a version of threshold deontology that instructs us to significantly favor deontological constraints over consequences at the highest level of abstraction on the four-stage sequence, we can start to interpret this at a mid-level such that its significance with respect to other, potentially-competing deontological notions is brought out. Once we know what kinds of mid-level possibilities for threshold deontology are open to us, we can begin the low-level investigations of threshold deontology, such as the question about where n is to be placed. Hence, I argue that my conception of moderate deontology, which provides a genuinely

moderate deontology by appealing to Rawlsian methodologies of principle interpretation, can easily accommodate Liao's response to Alexander, and hence can further show how, contra Smilansky, threshold deontology need not cease to be a genuinely deontological moral theory.

2.4. Non-lexical moderate deontology: answering Smilansky's challenge

The second section of this chapter attempted to expand and motivate Ellis's Legal Model for threshold deontology by fleshing it out in a Rawlsian framework of principle interpretation. In doing this, the hope is to show that even a loose method of interpretation of a general principle can result in a moral theory that retains the spirit of the general principle in a meaningful way. This is possible even if, as I have argued, the adjudication of thresholds on deontological constraints must be situational and will be imprecise. My discussion of Ellis, in essence, suggests a principled basis for replacing Smilansky's claim that deontology's distinctive feature is that it treats deontological constraints as lexically prior to consequences. By appealing to the Rawlsian four-stage sequence, I showed that deontologists can get by without lexical priority, making use of the imprecision inherent in the application of general principles downstream from their formulation. In this section, I further defend these claims by motivating the deontological character of this method. The problem I want to address arises for considerations of the following sort. For all I have said, nothing is to stop agents from using the four-stage sequence to develop a radically permissive threshold deontology. For instance, people may decide to set $n=5$, especially once we remember that, at the fourth stage of interpretation, anyone is conceived as engaged in interpretation of the general principle under the condition of full information. Hence, people may use the Rawlsian method to derive a low-level deontological principle which says that it is permissible to kill in the 5-to-1 Case discussed in the first section. If this is possible, and totally valid on the Rawlsian model for threshold deontology that I have proposed, then my

model appears too permissive. The four-stage sequence may be able to be applied in a perfectly sound way, but result in the general principle being applied in a way that is not deontology-preserving.

One possible answer to this problem is to argue that this kind of reasoning is just irrational for individuals who have accepted the non-lexical threshold deontology that I have proposed. To decide that it is permissible to kill in the 5-to-1 Case is just to think that deontological constraints are not to be significantly favored over consequences. If it is permissible to kill one as a means to saving five, then the constraint against this kind of killing just is not significant. So it's not clear how someone could reach this conclusion in the first place, if they understand the idea of deontology even a little bit. Yet I think we should not answer the problem this way, because it seems likely that bargaining down on deontological thresholds would be a significant phenomenon. If we set $n=10,000$, which is a number that some deontologists would find plausible as an a priori first pass at the matter, then it seems plausible that an objector might insist on a trivially lower threshold, such as $n=9,995$. It may be hard to see the relevant difference, from the standpoint of the interpretation of non-lexical threshold deontology, between $n=10,000$ and $n=9,995$. Now suppose that the objector wins out, and the threshold is set at $n=9,995$. A new objector might plausibly insist that $n=9,990$ is a plausible alternative to $n=9,995$. And so on. In this way, it is implausible to insist that setting n at a very low number is a priori irrational, given that thresholds will tend to be sorites-susceptible.¹

A different method for dealing with this seems necessary. I suggest that, instead of trying to deny outright that threshold deontologists can never set $n=5$, we undermine the reasons that

¹ Obviously, the question of how to deal with the Sorites paradox is a large one, and many solutions have been given (for an overview, see Hyde and Raffman 2018). Here, I specifically bracket all of this and pursue the question of how, morally, we ought to reason about practical Sorites paradoxes. This qualification is due to Gary Ostertag.

they might have for wanting this in the first place by getting clearer about what it means to institutionalize a moral theory. In doing this, I do not show that $n=5$ is an unacceptably low number, but that setting $n=5$ would require the threshold deontologist to shore up their moral theory in other regards, thereby imposing something like a mid-level principle that could disincentivize individuals from stretching threshold deontology that far.

Once we reach the fourth level of the four-stage sequence, mid-level principles govern the applicability of deontological thresholds with respect to other deontological values. As our discussion of Liao's response to Alexander showed, the second and third stages of the four-stage sequences could be marshaled to require deontological thresholds to be balanced against a person's rights to not be involved in harms in certain ways. At a mid-level, those involved in setting deontological thresholds might decide that individual attempts to press n downward are ruled out in a certain way. It might be held, for instance, that there are rules about how drastic changes to n might be: it may be simply decided that it is impermissible for people interpreting threshold deontology at the fourth stage to make more than a 5% change to the placement of the deontological threshold at any given time. In this way, some liberty to interpret threshold deontology at the lowest level might be accepted, while still constraining that liberty in a way that prevents people from bargaining down the threshold aggressively to get the result that killing is permissible in the 5-to-1 Case.

Importantly, this response introduces deontological considerations to explain why bargaining down the threshold would not be deontology-preserving. If mid-level principles were set up in this way, then radically bargaining down the value of n would require violating principles that themselves had a deontological foundation. This is because the mid-level principles are constructed through a public, quasi-democratic method, and hence are made by

moral legislators who are concerned to implement a plausible social system that instantiates moderate deontology.

To some more traditionally-minded moral theorists, this move will seem strange, and like it shifts the question. It might be thought that this procedure uses political fiat to set moral principles, rather than testing moral principles in the more traditional philosophical methods of normative ethics, such as intuition and reflective equilibrium. Yet it might be insisted that ethics and politics are not the same thing, and that there is no reason to think that appealing to a political notion would produce an answer to the ethical questions raised. In other words, this view might be construed as a kind of ethical skepticism, that says that ethical analysis past the first stage of application is just not possible except through non-ethical political analysis.

Although there is something to this challenge, I think the worry that it raises actually demonstrates a feature, not a bug, of the moderate deontology that I propose here. Suppose that it's right to say of this view that it shifts certain aspects of moral judgment to the political. To say that this undermines this view as a legitimate moral view assumes that there is no moral element to, or moral justification for, certain acts of political deliberation. But this is at least controversial, and at best irrelevant. It's not obvious that political deliberation, even at the highest level of abstraction, is non-ethical. Rawls's process for selecting the principles of justice are constructed in a way that brings moral considerations directly to bear. Similarly, insofar as persons lower in the four-stage sequence use those principles, as well as whatever else has been made available, to guide them in their various tasks, those persons make use of ethical analysis. In other words, we could see the insistence that the defense of deontology be separate from political notions such as democratic deliberation puts things exactly the wrong way. It is *because* a plausible Rawlsian political methodology is structured in a way that is answerable to ethical

considerations at bottom that the four-stage sequence understanding of the placement problem is plausible. It is also the explanation for why applying the four-stage sequence to define the contours of deontological thresholds is genuinely deontological. The sequence embodies the deontological assumptions that underlie the distinctively deontological considerations of Rawlsian liberalism. This is, of course, not an argument that we should accept moderate deontology, or any other moral theory, but just an argument that my approach to the placement problem is a genuine version of deontology, and hence is a response to Smilansky's challenge.

2.5. Rorty, political liberalism, and the method of moderate deontology

The four initial sections of this chapter connect the problems surrounding threshold deontology to a broader Rawlsian application procedure for deontology. The broader purposes of this dissertation, however, go far beyond this, toward the subject of Rortian reconstructions in normative ethics and metaethics.² Recall that the goal was to provide Rortian pragmatist reconstructions of moral theories that both answers to the need to endorse a moral theory as one's own and to do so in a way that respects the contingency of that person's vocabulary. To see how the Rawlsian considerations at play in this chapter connect to this broader project of demonstrating the possibility of pragmatist reconstructions, it is helpful to consider some of the claims made in Rorty's essay on Rawlsian liberalism, "The Priority of Democracy to Philosophy" (Rorty 1991d).

In that essay, Rorty sees Rawls's conception of liberalism as congenial to his general pragmatist framework, because both think that democracy, in political philosophy, is prior to philosophy. Rorty targets a conception of philosophy that takes its goal as stating metaphysical and epistemological truths about human nature that transcend local facts about humans located

² It is of tangential interest that Rawls (1980, p. 516) spoke highly of Dewey's pragmatism, insofar as both shared an interest in putting deontology in its proper place. The details of this don't matter much for our purposes.

where they actually are. Rorty's thesis in that essay is that he and Rawls are allied against philosophy, in this sense. Rawls, for instance, supposes that the principles that are derived at the first stage of application are principles that anyone would have reason to agree to, regardless of their broader philosophical beliefs about the nature of human beings and their views on the good. This is important because broader philosophical beliefs are never going to be the subject of the kind of agreement that needs to be in place if there is to be a legitimate conception of justice that can be applied in the four-stage sequence. Rorty, accepting Rawls's view on this matter, hence reads Rawls as holding that democracy is prior to philosophy in the sense that a democratic agreement can be set up and agreed upon totally independently of these broader philosophical questions, and can do so in a way that subordinates particular persons' philosophical conceptions to that democratic conception. In filling out the ideal of a democratic conception, Rorty appeals to Judith Shklar's liberalism of fear, which views the abhorrence of cruelty as a sort of minimal floor that all liberals can agree upon, with no reference to private philosophical argument (Shklar 1984, pp. 8-9, 32-34; Rorty 1989, p. 146).

In understanding how this project works, it is crucial that, for Rawls, even though the democratic procedures envisioned in his works can be the subject of reasonable agreement while not hinging on any of these substantive philosophical conceptions, which in later Rawlsian terminology are known as comprehensive doctrines (Rawls 2005, p. 12), one's endorsing a comprehensive doctrine should not preclude one from endorsing the democratic conception. Given the central role that diversity of viewpoint and toleration of diversity plays in a Rawlsian social contract, the democratic agreement reached has to be one that can accommodate many different substantive philosophical conceptions. If Rawls is right, then Kantians, devout Hindus, and utilitarians should all be able to agree on the democratic conception from within their

comprehensive doctrines. Again invoking later Rawlsian terminology (2005, p. 15), it must be possible for there to be an overlapping consensus on the democratic conception that accommodates all the comprehensive doctrines. Rorty does, however, read Rawls as committed to the claim that those whose philosophical conceptions require them to reject the idea of the overlapping consensus, which place them necessarily outside of it, cannot be a part of the democratic conception (Rorty 1991d, pp. 187-189). It is an open question what we should do with these individuals, given that toleration comes from within a democratic conception, and hence seems to exclude them.

There is now a problem facing us. Initially, I'm inclined to say that Rorty would count moderate deontology as a private moral theory, and hence as a philosophical conception. Yet the Rawlsian procedure that I have invoked to provide a plausible account of moderate deontology is substantially public, a crucial aspect of Rawls's free-standing, non-philosophical democratic conception. This might be taken to show that, from a pragmatist perspective like the one that Rorty endorses, I have made a kind of category mistake in trying to make moderate deontology more legalistic, procedural, and public. The account would no longer belong clearly to the public/democratic or to the private/philosophical. The distinction between democracy and philosophy strongly suggests a distinction between the public and the private: philosophical conceptions are relegated as a matter of individual or group conviction, while democracy, along with its requirements for radical toleration and accommodation, is supposed to form the available-to-all normative substance that conditions the possibility of a society. My project faces the further question about whether non-lexical moderate deontology can avoid being philosophy, in the pejorative sense endorsed by Rorty in "The Priority of Democracy to Philosophy". It's not clear why a first-level deontological principle that requires us to grant some priority to

deontological considerations should be the object of democratic consensus, given that democratic societies must speak to the concerns of all citizens, and that utilitarians number among those citizens. So there is now the worry that, in setting up this system the way I have, I have not succeeded in providing a conception of threshold deontology that is an adequate reconstruction of deontology, better tailored to the facts of pluralism and disagreement about moral theories, but instead simply provided one private moral theory among others.

I think we should resist this line of argument, because it fails to appreciate the value in modifying deontology in this way. Suppose we think that democracy is prior to philosophy. Viewing that as an objection to modifying a private moral theory to make it closer in form to the form of a conception that could have a place in a democratic consensus is to effectively assume that democracy has to be totally abstracted from philosophical conceptions in all cases. But as both Rorty and Rawls grant, this is unrealistic, and to ignore the historicist and culturally-mediated character of persons (Rorty 1991d, pp. 180-1). We all find ourselves as a part of an already-existing tradition. Hence, it is a crucial project, in order to have actual public discourse with a diverse group of individuals, to find aspects of our own, private philosophical conceptions that can be tweaked and revised so as to enable that discourse. Rawls, in other words, as well as the conception of moderate deontology I have outlined here, can be read through a Rortian lens as offering a methodology for moving our comprehensive doctrines into the public sphere, by applying the more general conception which might indeed be private to application schemes that render them more particular and more sensitive to the rhetorical demands that diversity places on moral agents.

Incorporating a democratic consideration into one's private philosophical view is nothing close to a vice, or even less a category mistake. It is merely a way to recognize the contingency

of the constituents of one's moral vocabulary. This effect, I think, is greatly freeing. It shows that embracing contingency, and allowing it to filter down and inform our own moral theorizing, can produce the benefit of incentivizing us to make more flexible, nuanced theories. I should also mention in passing that the possibility of pursuing private moral theory in this way raises a serious pragmatist question about Rorty's use of the public/private distinction in *Contingency, Irony, and Solidarity*. Dividing the moral world up in this way makes sense to Rorty because he views private moral theory through a pragmatist lens that is inherently skeptical about the value of private morality (Rorty 1989, p. xiv). Philosophy in the pejorative sense is incompatible with the burdens placed on citizens in a democratic society that itself demands legitimation. But pragmatism provides us the tools and motivations to shape our philosophical conceptions in light of the democratic conceptions that we find ourselves a part of. Citizens interested in engaging with the public considerations of a democratic society can do so in a way that is consistent with their own private values, at least sometimes, because there is this methodology of reconstructing one's own moral vocabulary in light of the demands that democratic society places on citizens.

A brief conclusion. This chapter deals with the relationship between moderate deontology, the Rawlsian four-stage sequence, and Rorty's reading of the latter in terms of the private/public distinction. The idea is that a suitably moderate deontology, augmented by the Rawlsian structure to explain how deontological considerations can be prioritized without endorsing lexical priority, is acceptable on a Rortian conception of the liberal ironist and the various vocabularies that they must reconcile. A moderate deontologist should be a better-engaged liberal ironist, involved in the difficult and ongoing task of sensitizing one's own philosophical conceptions to public considerations, and vice versa. A liberal ironist can be a moderate deontologist, and can do so in a way that is predicted by and appropriately responsible

to the aims of a pragmatic liberal ironist. It also follows that threshold deontology cannot be meaningfully applied without mid-level principles, and hence that any attempt to reduce moderate deontology to threshold deontology is bound to be incomplete. Hence, this chapter has both made a contribution to the technical literature on moderate deontology and made a broader point about how pragmatists and liberal ironists can relate endorse, use, and defend deontology with a genuine and broad appreciation of its internal nuances.

INTERLUDE: From Pragmatist Normative Ethics to Pragmatist Metaethics

Recall that Chapters 1 and 2 were pitched as a way to show that Rorty was wrong in his insistence that pragmatism and liberal ironism must be skeptical of the idea of a moral theory, because moral theories improperly aim to provide a final vocabulary for ethical inquiry. Rorty is wrong about this because the Rortian pragmatist just needs to do some work to make moral theory practical without final vocabularies. It is indeed possible to hold fast to a conviction in a moral theory in a meaningful way. It is also possible to do so in a way that makes it possible to actively debate our moral theories, and to restructure them in response to criticism. In short, the project of moral theory can be vindicated even on pragmatist assumptions about the difficulty of doing normative ethics. This is because moral theories can be multi-leveled, containing complex internal structures that allow them the plasticity to pivot to the situational demands placed on them, enabling the reconstruction of moral philosophy. I take this as a significant point in favor of Rortianism.

However, a serious problem still looms over this pragmatist reconstruction of normative ethics. Many have thought that the positive project of normative ethics requires a realist metaethical theory which provides at the very least a realist account of moral methodology that gives moral claims a strongly realist epistemology and an account of moral semantics that makes sense of how moral sentences refer to ontologically-real moral properties. Again recall Parfit's claim, cited in the Introduction and discussed in Chapter 1, that aiming at truth, in an ambitious sense comparable to the sense of truth that is found in mathematics, guides the project of moral theory. Without having been able to provide this, instead opting to take a pluralistic attitude toward theories that aims to let various moral theories live together, it might be said that my arguments do not really vindicate moral theory. The arguments in Chapters 1 and 2 give us some

indirect reason to think that Parfit, and others who agree with him, are wrong about this. If they were right about this, then the perspective present in those chapters would not have been able to be made to look plausible. But I think it would be better to attack the issue straightforwardly. Hence, Chapters 3 and 4 aim to directly show that Rorty's pragmatism, with some modifications, is well-placed to do moral theory while remaining separated from realist metaethics. The method for doing this is specifically pragmatist in spirit: it surveys various realist and antirealist metaethical positions, and aims to take their influences on board and to construct a distinctively pragmatist approach to metaethics that emphasizes the possibility of talking the talk of certain moral realist commitments, like the concept of moral objectivity and the method of intuitions, in a way that legitimately answers the concerns of those who take those commitments seriously. In other words, the following two chapters provide a pragmatist reconstruction of metaethics.

CHAPTER 3: Intuitions in Ethics From a Pragmatist Perspective

3.1. Introduction: pragmatist skepticism about intuitions

Historically, pragmatists have been highly skeptical about the role of intuition in philosophy. This is because intuition is often linked with various heavy-handed philosophical notions such as the quest for a priori knowledge, foundationalism, or non-naturalism, notoriously hard to make sense of on the pragmatist framework. C. S. Peirce, for example, argues that intuition, understood as “a cognition not determined by a previous cognition of the same object, and therefore so determined by something out of the consciousness” (Peirce 1992, p. 11), even on the supposition that it is possible, is methodologically infirm, in that we can’t know anything about whether some cognition is intuitive or not (Peirce 1992, p. 20). Hence, it is unclear how intuition could matter to a pragmatist, assuming that we need to know stuff about our intuitions in order to use our intuitions as a tool to aid inquiry. Although Quine said little about intuition, and although Quine was happy to appeal to intuition in his arguments, for example in his discussion of natural kinds (Quine 1969b, p. 125), it seems plausible to say that Quine would reject the granting of a strong epistemological role to intuition insofar as that granting presupposed some account of the a priori or of analyticity. And of course Rorty, in *Philosophy and the Mirror of Nature*, voices skepticism about the possibility of intuition’s being able to prove dualist claims in the philosophy of mind insofar as intuition is understood on a Wittgensteinian tack as a mere certain kind of step in a language-game (Rorty 1979, p. 34), and hence as having no a priori role in philosophy. Intuition, on his view, is historically-conditioned down to the bottom. Cornel West concurs with Rorty, specifically noting Dewey’s rejection of the a prioricity of all forms of evidence, intuition presumably among them (West 1989, p. 110).

Yet, with all that has been said, and with all the intellectual weight behind these claims, I will argue in this chapter that this approach to intuition in epistemology and ethics is wrong. Because we can sever the link between intuition and the metaphilosophically-bloated epistemological notions often associated with it, it is possible to account for intuition in a way that makes sense of the crucial role that intuition plays in our practices. Indeed, then, we see that intuition not only survives an anti-realist critique, but does so for reasons that are deeply related to the core of pragmatism's aim to reconstruct philosophical notions by strategically revising them in light of the failure of the traditional conceptions to do what they need to do. It is because our actual, reasonably-accepted practices demand this role for intuition that we can justifiably, sometimes, rely on intuitions in argument.

With that said, it's time to begin. The first task is explicating more clearly the concept of intuition, insofar as Peirce's notion has been surpassed by a century and a half of philosophical work. Hence, in the next section, I show that there are two notions of intuition, which we might call the rationalist and the empiricist conceptions, that are nonetheless unified by their attempt to play a certain methodological role in moral and epistemic discourse. Then, in the successive sections, I will argue that neither the rationalist nor the empiricist conception can actually play that role. Rationalists, it is argued, cannot provide an account of a methodology for disagreement resolution for intuitions, in virtue of their overly demanding notion of what our practice of eliciting intuitions aims at. Empiricists, in contrast, due to their more relaxed conception of the aims of intuition, can make sense of the methodological efficacy of intuition, but cannot make sense of its non-emotional surface form and cannot make sense of the rational revisability of our intuitions. I then proceed to argue that naturalism about intuitions can both avoid all of these problems and be shown to comport with pragmatism: its aims are properly in step with its

methodological possibilities, and can make sense of the surface form and rational revisability of intuition.

3.2. Some views of intuitions

The locus classicus of the empiricist conception of intuition is Jonathan Haidt's "The Emotional Dog and its Rational Tail", in which empirical results, notably the phenomenon of moral dumbfounding, are taken to show that intuitions about ethics are primarily affective, that is, driven by emotional gut reactions to our consideration of ethical cases. Certain gut responses to cases are famously immune to reasoning; it is very difficult to use rational argument to get a person to give up their intuition that for example incest is morally wrong.

Haidt's discussion in that paper draws out a number of key features of intuitions, but in this paper I will focus on Haidt's definition of intuition, which on his view is "the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of searching, weighing evidence, or inferring a conclusion" (2001, p. 818). In defining intuition this way, Haidt assumes something like the dual process theory of moral judgment, according to which moral judgments are either the result of intuition, conceived of as analogous to the 'automatic' mode of a camera, or reasoning, conceived of as analogous to the 'manual' mode of a camera (Greene 2014, p. 696).

Extrapolating a bit from that definition, it seems fair to say that, on Haidt's view, intuitions are immediate (see also Sinnott-Armstrong, Young, and Cushman 2010, p. 246). An intuition is immediate just in case it "[does] not arise from any process that goes through intermediate steps of conscious reasoning" (Sinnott-Armstrong, Young, and Cushman 2010, pp. 246-7). Although it is an interesting empirical fact that many of the judgments we have that arise

from intuition are immediate, it is also important to keep this in mind, because the rationalist conception of intuition is also beholden to this fact, for similar reasons, but spins this fact in a different kind of direction. We will distinguish *causal* immediacy from *epistemic* immediacy. Causal immediacy is the property analogous to what Jonathan Dancy calls a “presentational state” according to which intuitions are, as a matter of psychological fact, “fundamentally nonvoluntary, being things that just happen to us, come upon us unasked” (Dancy 2014, p. 792). Although Haidt doesn’t use this term, the idea of a presentational state is endemic to his definition of intuition, and is quite nearly built into Greene’s idea of automatic, intuitive reasoning as reflex-like (Greene 2014, p. 696).

The empiricist conception of intuition has been taken to have skeptical consequences. Suppose that the social intuitionist’s conception of intuition is correct when it claims that most moral judgments are based on intuition, the way the social intuitionist conceives of it. If moral reasons have to be objective and have to be reasons-responsive, then the social intuitionist’s conception of intuitions cannot account for moral reasons (Juzaszek 2016, p. 64). All it can give us is, to borrow Haidt’s term quoted above, affective valence, and nothing that can play the epistemic role of justifying a moral judgment. Since most metaethicists and normative ethicists accept the empirical facts that motivate the empiricist conception of intuition, it has become incumbent to show how intuitions can play this epistemic role, in the face of these challenges. How might an intuition play an epistemic role, given that intuitions are immediate and not conscious in their phenomenology? The answers we are about to give to this question will allow us to delineate and propose the rationalist conception of intuitions, and to distinguish it from the empiricist conception.

The rationalist conception of intuitions holds that intuitions can sometimes be epistemically immediate, meaning that the presentational and immediate status of intuitions sometimes itself has an epistemic valence. Although this position has been held by many (Shafer-Landau 2003, p. 238; Audi 2004, p. 15; Parfit 2011b, pp. 502-3; Huemer 2005, p. 102), and we are often comfortable appealing to our intuitions in moral contexts where the goal is to support or make a knowledge claim, it can be hard to see the reasons that someone might hold such a view. To account for this, I want to explain two possible ways to account for epistemic immediacy.

The first answer to this question requires us to delve into the idea of *intuition as intellectual seeming*. The question above asks how intuitions can play an epistemic role, given that they are immediate, not conscious, and hence cannot be the product of explicit reasoning. The answer to this is to give an account of how people can form judgments in an epistemically reliable way without requiring actual reasoning. But this, perhaps, is not so hard to understand. Consider a commonsensical account of how perception works. Our visual perception is quite epistemically robust, even without being consciously controlled or explicitly undertaken. Under normal conditions, when I see the white wall in front of me, I would be epistemically justified to a high degree of confidence that there is a white wall in front of me. We might also say that, without significant evidence that there is not a white wall there and that my perception is not functioning correctly, there is a presumption in favor of believing that my judgments are correct. This presumption is named Phenomenal Conservatism by Michael Huemer, who explains the epistemic status of intuition by analogy to the epistemic status of perception (2005, p. 99-105). To bring out more specifically what this involves, take George Bealer's formulation of the idea of rational intuition (Bealer 1998, pp. 208, 216-7), according to which intuitions are intellectual

seemings that are strongly modally connected to the truth (see Pust 2000, p. 39, for a similar view). When a person is in ideal cognitive conditions, its seeming to that person, intellectually, that p is strong evidence that p. Bealer introduces terms like “ideal cognitive conditions” and “strong modal connection” because he wants to account for both the fact that people’s intuitions are sometimes wrong and the fact that, even in spite of this, we have a strong justification for relying on intuitions. This grants a presumption in favor of accepting our intuitions.

Ernest Sosa has questioned the intuition/perception analogy, and defended a rationalist conception of intuition that grounds the epistemic role of intuition in terms of epistemic virtue. For Sosa, intuition and perception are importantly disanalogous in that there are no experiences present in the act of intuition that play the role of sensory experience in perception: “What *intuitive* justification lacks is any correlate of the visual sensory experience beyond one’s conscious entertaining of the propositional content, something that distinctively exerts a thereby justified attraction to assent” (2007, p. 55). Yet Sosa accepts that intuitions are intellectual seemings (2007, p. 54). He hence explains the epistemic role of intuitions as issuing from epistemic competences. Sosa writes that “a competence is a disposition, one with a basis resident in the competent agent, one that would in appropriately normal conditions ensure (or make highly likely) the success of any relevant performance issued by it” (2007, p. 29). To have a justification on the basis of an intuition, then, is to have a certain intellectual seeming that is competent. Sosa avoids the problem of the intuition/perception analogy by showing that intuition and perception are both epistemic competences, albeit in different modes, in that they are different sorts of intellectual activities (c.f. the discussion of colorblindness, Sosa 2007, p. 59).

Considering Bealer’s and Sosa’s accounts in tandem is helpful because it shows us that the rationalist account of intuitions endorses a methodological claim and meets the justification

challenge of empiricism. We might take Bealer as specifying the concept of intuition as appealing to intellectual seemings to explain its epistemic role, and Sosa as working out an account of what criterion we might appeal to in order to explain what, in a more commonsensical way, an intuition really is.

So far, Bealer and Sosa have established the possibility of a rationalist epistemology of intuitions, but let's go deeper into how something like a method of intellectual seemings is operationalized. Robert Audi (2004, pp. 45ff) proposes a method for doing just this by appealing to what might be called the method of holistic reflection. Audi thinks that we have the capacity to allow our judgment to be influenced by broad-scale, holistic consideration of the relevant features of the situation in which we are to form our judgment. Audi likens the experience to that of judging the quality of a painting from considering its various physical properties in this manner. Appealing to something quite like Bealer's ideal cognitive conditions, Audi's idea is that in at least some cases, forming our intuitions in this way is warranted. Hence, Audi can be read as expanding Sosa's conception of intuition as intellectual competence, by specifying a method by which reflection can be made competent, and hence able to play an epistemic role.

Earlier, I pitched this discussion as one way to answer the question regarding how an intuition could play an epistemic role, given that intuitions are immediate and not conscious in their phenomenology. One way intuitions could play this epistemic role is by being an intellectual seeming. A second way, that I will now address, is by being granted a certain status for the purpose of normative reasoning. On this view, intuitions are the data points that normative ethical theorizing responds to. I call this the *intuition as data point* view. This view is discussed in the first section of Shelly Kagan's paper "Thinking About Cases." Kagan writes that, in the course of trying to construct a moral theory, "[i]ntuitive reactions to cases—real or

imagined—are carefully noted, and then appealed to as providing reason to accept (or reject) various claims” (Kagan 2001, p. 44). Kagan then proceeds to claim that intuitions are taken to be evidence for or against moral theories, but not absolute evidence, as we typically treat our intuitions as possibly revisable or rejectable. This is even more crucial when we note that we sometimes have intuitions about our principles as well, and that sometimes our intuitions about cases and intuitions about principles will pull in different directions.

In explaining the intuition as data point view, it is helpful to consider some arguments given in Frances Kamm’s *Intricate Ethics* (2007), which shows this way of thinking about intuitions out in full force. According to the intuitions as data points view, it is the aim of a moral theory to account for our intuitions. Intuitions are the data which check our moral theories, and which can be used to press for changes in moral theories. Consider Kamm’s discussion of the Doctrine of Double Effect (DDE), in particular the way that it is used to provide a solution to the Trolley Problem. Kamm’s problem is to provide a theory that harmonizes judgments that are commonly had about certain trolley cases. Three example cases are necessary to show how this works (Kamm 2007, p. 92):

Trolley: “[T]he standard Trolley Case involves, let us say, five people on a track toward which an out-of-control trolley is heading. If it continues, it will kill them. However, if we redirect it away from them, it will go down another track where it will kill someone who is unable to be removed.”

Bystander: “Another way to stop the trolley from hitting the five is to push an innocent bystander in front of it. Its hitting him stops it, but we foresee that he will die as a consequence.”

Loop: “[E]verything is as it is in the Trolley Case, except that the track to which we can redirect the trolley away from the five loops back toward them. Were it not for the presence of the one person on the side track whose being hit stops the trolley, it would go around to kill the five shortly.”

According to the standard story about these cases, it is permissible to act in Trolley, impermissible to act in Bystander, and permissible to act in Loop. Yet many have noted that this pattern of reactions is strange; explaining it, and other patterns of reaction that include different intuitions, is known as the Trolley Problem (Thomson 1985). First, it is unclear why the means matter morally in the first place; for consequentialists, it seems weird to insist that there is an actual moral difference between Trolley and Bystander in the first place. Second, and more importantly for our purposes, once we take nonconsequentialism on board and grant that there is a relevant difference between Trolley and Bystander, it is still unclear why our reactions in Bystander and Loop should differ. In both cases, the presence of the person who blocks the trolley is causally intended as a means to the saving of the five. If the person's body were not available, then the action wouldn't be possible. So why would a principle like the DDE, which says that it is not permissible to causally intend harms as a means to the saving of lives and is often touted as a resolution to the Trolley Problem, accomplish its purpose here? Kamm's solution is to amend the DDE to the Doctrine of Triple Effect (DTE), which distinguishes between causally intending a harm *because* it will bring about an effect, and causally intending a harm *in order to* bring about an effect. Kamm says that DTE gets the right answer in Bystander and Loop, in terms of the standard story above, because in Loop, although the harm is causally intended, it is causally intended because it will bring about an effect, whereas in Bystander, the harm is causally intended in order to bring about an effect.

Given the enormous complexity of Kamm's argument, it is fortunate that the details of it don't matter here, because we're interested in the methodology that she employs. And although in the course of her argument, she doesn't use the word 'intuition' to refer to the data that result in the pattern of judgment that she wants to account for, it is clear from other passages in

Intricate Ethics that the relevant data are intuitions. She writes, for instance, when discussing the broader nonconsequentialist theory that results from her discussion of the DTE: “in this section I shall propose components of the theory followed by a discussion of cases in which intuitions can be accounted for by that component and cases in which intuitions call for further refinements of the theory” (Kamm 2007, p. 138). Correspondingly, she dedicates a later chapter of her book to a defense of intuitions in the face of empiricist worries about intuition. She is specifically concerned to explain how intuition can be reliable in the face of concerns about framing effects in classic studies by Daniel Kahneman and Amos Tversky. Walter Sinnott-Armstrong, in his discussion of the objection to intuition from framing effects, writes that framing effects are “effects of wording and context on moral belief” (Sinnott-Armstrong 2008, p. 52). Sinnott-Armstrong continues with an example: “Imagine that Joseph would believe that Marion is fast if he is told that she ran one hundred meters in ten seconds, but he would not believe that she is fast (and would believe that she is not fast and is slow) if he is told that it took her ten seconds to run one hundred meters (or that it took her ten thousand milliseconds to run one hundred meters)” (Sinnott-Armstrong 2008, p. 52). In this example, Joseph’s intuition is unreliable, because the order of presentation of the words should not be part of what justifies this kind of judgment. But supposing that intuitions in ethics were subject to these same kinds of framing effects, that would show that those intuitions were unreliable, too. Sinnott-Armstrong (2008, p. 68) uses this conclusion to argue that our intuitions need inferential confirmation. Kamm responds to this sort of challenge by arguing that, if we can know enough about framing effects to point them out in people’s judgments, then we can re-frame cases so as to control for those effects, thereby producing reliable intuitions. Consider the following:

[E]xperimental subjects give different responses about the necessity of a health-care policy to deal with a coming [viral epidemic], depending on whether we say either that

(a) without the policy, 400 of 600 people will lose their lives, or (b) without the policy, we can only save 200 of 600 people. In description (a), the baseline suggested by the phrasing is a state in which people are now well but face getting worse; the baseline suggested by the phrasing in description (b) is the near-death state people will be in if there is no intervention, but from which there can be improvement. Subjects think that it is worse if people lose their lives than if they do not gain them, and they are more averse to a policy in which people lose their lives than one in which the same number are not saved (Kamm 2007, pp. 424-5).

But suppose, with Sinnott-Armstrong, that the framing of the baseline is morally irrelevant; we can still say that intuitions can be reliable if the baseline is always framed in the correct way.

Hence, Kamm proposes that our intuitions can be trusted when they have been framed so as to avoid framing effects. In judging that, for example, order is an irrelevant framing effect, a theory of what effects are relevant is assumed: “framing effects are supposed to lead to different responses that cannot be reasonably justified on account of the frame, that is, the frame must introduce what is seen, at least on reflection, to be a difference that should be irrelevant to different responses” (Kamm 2007, p. 439). Kamm then argues that cases where researchers like Kahneman and Tversky assume that a framing effect is irrelevant might actually be cases where the intuition tells in favor of the framing effect’s relevance, and not against the theory that assumes that the framing effect is irrelevant. For example, an exemption from paying a penalty for doing an action that is the same size, in dollar amount, as a surcharge for performing that action under an alternate scheme of enforcement is supposed by some to be an irrelevant difference, but it might actually be morally relevant if the surcharge is conceived as depriving the actor of something to which they have a right (Kamm 2007, p. 440).

What arises from Kamm’s discussion of these matters is that the presence of framing effects cannot undermine appeals to intuition to do ethics. This is because the relevance of framing effects cannot be known in advance of the theories that those intuitions constitute the data of, because we can argue about the significance of framing effects, and therefore how big of

a problem a framing effect is for an intuition. Ethics is a complex negotiation of our intuitions about cases and the theories that inspire those intuitions, and how they can be made to work together. In Kamm's theory, the relevance of an intuition is not that it is an intellectual seeming that carries with it some epistemic role, but merely that it constitutes the data that allow us to derive and check our moral theories. Simply figure out which framing effects are justified in the case, and we know when to write off certain intuitions.

I could keep going, but the point about intuitions as data points is hopefully clear. Now that we have both of these rationalist notions of intuition in ethics on the table, the problem is to state a methodology for intuition in ethics. Rationalists typically think that intuition can play an epistemic role in ethics because intuition can be given a methodology for investigation, somewhat analogous to the role played by the scientific method in scientific investigation, specified by the intuition as intellectual seeming view and the intuition as data point view. As a matter of fact, these methods can be made to work together, as I will discuss a bit below. For now, though, it is helpful to raise the question as to whether, despite appearances, this discussion shows that the rationalist conception of intuitions can uniquely play these methodological roles. Another way to ask this question is to ask whether an empiricist can make sense of intuitions as intellectual seemings or intuitions as data points. If it turned out that empiricist conceptions of intuitions could make the case that social intuitions, on Haidt's line, could play one or both of these epistemic roles, it would show that stronger rationalist claims about intuitions were unnecessary.

Before moving to the next section, I want to take stock of what has been said here. I have elaborated the empiricist conception of intuition and the rationalist conception of intuition as differing on the epistemic status of moral presentational states, and suggested that the rationalist

conception is distinguished by its holding that intuitions can be epistemically valenced, in contrast to the empiricist conception, which holds that the valence is merely causal. I then elaborated two methodological aspects of the rationalist conception that are supposed to explain how intuitions can be epistemically valenced: intuitions are intellectual seemings, and intuitions are data points in moral theorizing. Yet in the first section of the paper, I framed my project as one of arguing that there is something like a pragmatist conception of intuitions that can harmonize these accounts and still play the methodological role we expect of intuition in ethics. With that said, then, the task now becomes showing that a pragmatist conception of intuitions would have methodological benefits over both the rationalist conception and the empiricist conception; neither the rationalist conception nor the empiricist conception can make good on the methodological role needed for intuitions alone. By bringing these accounts closer together, we can illuminate crucial aspects of moral deliberation that would otherwise have been ignored.

3.3. A methodological problem with the empiricist conception of intuitions

If providing a methodology for intuitions requires that we make good on the idea of intuitions as intellectual seemings, then the empiricist conception of intuitions surely cannot provide a methodology for intuitions, and does not even want to. The empiricist conception, in its explanation of intuitions as causally-valenced, emotionally-mediated presentational states, can explain why our intuitions are deeply-held and non-negotiable for particular individuals, but it cannot explain why that fact can justify our intuition's having anything like epistemic privilege for the agent. In other words, it cannot explain the epistemic status of intellectual seemings. However, it does seem like the empiricist conception of intuitions can make perfectly good sense of the methodological role of intuitions as data points for moral inquiry. We might take intuition, on this view, as simply signaling for an agent that some consideration ought to be accorded a

more fundamental role in discourse than some other consideration, defeasibly, given other aims and values that they hold, and this can be made sense of without requiring us to take a stronger view about the epistemic properties of the intuition. It is as if we treat intuitions as data points because they are more centrally-located for us, rather than because they are based on some a priori foundation, as a rationalist perspective on the methodological role of intuition as intellectual seeming seems to require. Hence, although this methodological role is not properly epistemological, as the rationalist thinks it needs to be, the methodological role is explanatory of why it makes sense for us to engage in particular kinds of moral discourse from the empiricist perspective.

With these arguments, I take it as clearly shown that intuitions, on the empiricist's conception of intuitions, can play a methodological role, and hence that a rationalist critique of the empiricist conception of intuition that alleges that empiricist intuitions cannot play any methodological role in discourse would be unfounded. A more nuanced rationalist critique, however, would argue that even if intuitions can play some methodological role for the empiricist conception, it cannot play the right epistemic role. In alleging that the methodological role for intuitions arises because certain intuitions are taken by the subject to be data points for moral inquiry, there needs to be an epistemology of which data points we take, because without that, there might simply be differing conceptions of the normative data points. An epistemic role for intuitions, cashed out as intuition as intellectual seeming, would then have to provide a method for deciding between normative data points, or as reconciling disagreements between various intuitional data points. Yet while granting intuitions as having merely causal valence, the empiricist cannot obviously make good on this. Hence, the empiricist conception has not shown

that it can provide a methodological role for intuitions that is robust enough to decide between normative data points.

Here, the rationalist will claim advantage, because the intellectual seemings account can apparently provide such a method. Audi, for instance, thinks that holistic reflection can undermine some intuitions, and Kamm thinks that intuitions can be reframed. In the next section, however, I argue that the intellectual seemings account actually does not help the rationalist, or that even on the assumption that it is possible for appeal to intellectual seeming to help provide a method, that the range of application of that method is so narrow that the insight is uninteresting. As it turns out, the rationalist wants more than any method of intuitions can achieve.

3.4. Methodological problem with the rationalist conception of intuitions

In the previous section, I argued that the empiricist's conception of intuitions cannot answer to any meaningful success standard. Hence, it was reasonable to conclude that the empiricist's conception cannot play enough of a methodological role to make it worth having in the first place. This provides some reason to explore in more detail the rationalist's conception of intuition, to show whether it is possible for the rationalist's conception to play that role. Initially, it would seem as if the rationalist is in a better position, because the rationalist can employ a stronger methodology to account for the epistemic valence of some intuitions. As it turns out, though, the rationalist is in no better position with respect to the epistemic valence standard. Although the rationalist can explain how a very strong interpretation of the method of intuitions can be used to do productive moral work, and hence do something like satisfy the epistemic valence standard, the rationalist does this at the cost of narrowing the scope of application of the method, making the method of intuitions usable only in a radically narrow set of cases, where people are already committed to a philosophical notion of consistency and a philosophical

methodology. The majority of people, on the rationalist's conception of the method of intuitions and how it can be used to do work, would be left with the empiricist approach.

The argument turns on two crucial ideas, which I will introduce now. First, it casts the problem of substantiating the methodological role of intuitions as a problem about the calibration of our intuitions. The so-called calibration problem comes from a paper by Robert Cummins (1998), who argues that the epistemic role of intuitions requires that intuition be verified independently of the faculty of intuitions. This verification could take the form of either verifying the faculty of intuition itself, or by verifying particular intuitions, one-by-one, by some other means. So far, so good. The problem, however, is that neither of these things can actually be done. We cannot verify our intuitions one-by-one, because if we are able to verify our intuitions by some other method of knowledge-acquisition, then the intuition itself did no epistemic work. The intuition was merely a signal for some idea that was then justified. That means that it was actually the other, non-intuitive thing, that has epistemic valence. On the other hand, we also cannot verify the faculty of intuition itself, for reasons often noted by philosophical skeptics, who plausibly hold that there is no methodologically-independent way of verifying a method of knowledge acquisition. We rely on perception to give us knowledge, for example, without any argument that our perception is reliable. If both arms of the calibration problem can be sustained, then this means that we cannot account for the epistemic valence of our intuitions, and the rationalist's conception of intuition must be rejected.

This framing of the calibration suggests two possible rationalist responses. First, the rationalist may simply rest content with their inability to calibrate their intuitions. No attempt, on this view, is made to verify particular intuitions. Rationalists simply rest assured that intuition as a faculty rests on as firm a foundation as any of our other faculties. This response rejects the

implicit premise in the calibration problem which says that we need to verify any of our epistemic faculties. We are perfectly justified in forming beliefs on the basis of our perception, at least in many cases, just because we have no choice but to sometimes rely on perception. To demand verification for any of our epistemic faculties creates a calibration problem for that faculty. The structure of this argument is basically to find companions in guilt: on the supposition that any demand to verify an epistemic faculty leads to skepticism about that faculty, because independent verification of a faculty is not possible, then demanding of all of our epistemic faculties that they be verified leads to global skepticism; and hence, we ought to reject the supposition that our epistemic faculties demand verification in the first place (Huemer 2005, pp. 108-9). A second rationalist response to the calibration problem accepts this supposition that we should verify our epistemic faculties, and holds that verification of intuition is possible. These rationalists employ holistic reflection to verify intuitions when they are used as data points. We can see Kamm's method here, for instance, in trying to show that some intuitions, insofar as they are determined by framing effects, ought to be rejected. The problematic framing effects are discovered by reflecting on the seeming moral relevance of those intuitions framing effects. Order effects, for example, as discussed before, are irrelevant, and intuitions that service to them ought to be rejected. S. Matthew Liao, by contrast, argues that we can provide an active demonstration of the possibility of calibrating intuition is by attending to the numerous examples where philosophers have actually used reflection and reasoning to calibrate their intuitions (Liao 2007, pp. 257-60). On this view, we solve the calibration problem by actually calibrating our intuitions, through the use of intuition as epistemic faculty. In other words, we can calibrate our individual intuitions, so long as we operate the rationalist conception of intuitions carefully enough.

The second crucial idea on which my arguments against the rationalist conception of intuitions turn is the claim that a rationalist conception of intuitions ought to provide a method for solving disagreements about our intuitions. This was also a theme in my criticism of the empiricist conception of intuitions, so it will pay to be more explicit about it. Although disagreement is one of the most heavily discussed topics in metaethics, most of that literature has been focused on problems in establishing the logical possibility of disagreement, given the complexities of positions argued for in the literature on moral semantics, or on connecting disagreement to issues regarding peer disagreement, and the epistemology of believing an agent's testimony (for an overview, see Frances and Matheson (2019)). The problem of disagreement that I discuss here is significantly different. Here, I discuss moral disagreement as a practical problem that real agents confront, and which might be aided by the contributions of philosophers.

Putting together these two ideas, the rationalist is confronted with a thorny question: Is it possible to employ a method for calibrating our intuitions that can reliably settle interpersonal disagreements about intuitions? It seems like rationalists should be able to answer 'yes' to this question, on pain of providing an error theory about the epistemology of intuitions. But it seems like the rationalist actually has to answer 'no'. I argue that this fact provides reason to reject the rationalist's conception of intuitions, as it is stated by those rationalists.

Although the resolution of disagreements is hard, on any methodology, this is particularly hard for the rationalist. Huemer (2005, p. 142) notes that rationalism entails that many moral disagreements are resolvable, in that the method of intuitions can be invoked to help parties come to agreement about who was right and who was wrong, but he specifically claims that the method need not work often, or nearly ever at all. The method is about establishing the

possibility of rational disagreement and its resolution, but the fact that people are unable to do this in practice is no issue for the method. We don't after all, think the fact of inability to resolve disagreements in science is evidence against the scientific method. In this, of course, is a kind of tacit admission that the method itself is infirm, and an attempt to reconceive the aim of ethics in face of that fact.

Nonetheless, it might be said that Huemer and those who think like him box themselves into a corner, precisely because of the way that they view success in the resolution of disagreements as a kind of quasi-scientific analysis. On such a view, the resolution of disagreements looks quite like a thing which is logically possible but practically unlikely precisely because moral beliefs are deeply held, hence placing serious cognitive demands on agents who want to change them in others. There is, in other words, a rational persuasion model of moral disagreement resolution. I propose to grant this, and to grant that from this, important claims about rationalism as a method for resolving disagreements about intuitions follow. At this point, it will be useful to remember that this is up for discussion because the empiricist conception of intuitions, it was seen, could not answer to crucial challenges that arose given the aim of explaining the epistemic status of intuitions. The empiricist conception told us that we should treat intuitions as things merely held constant for the sake of argument, which themselves had no epistemic valence on their own. This was seen to be unsatisfying, and to render the idea of a method of persuasion unintelligible. Rationalism about intuitions, and its stronger standard of epistemic valence, in contrast, could at least make sense of this notion of intuitions. But by doing this, we see that the rationalist has replaced the problem of the lack of epistemic valence with the problem of too much epistemic valence, by setting up a standard for moral success that is just too strong to answer to most cases. To take Huemer at his word, disagreement resolution

is extremely cognitively demanding. The way I see it, then, we have three choices in the face of this: either we can insist that disagreements are generally not resolved, or we simply retreat to the empiricist notion that there is no such thing as epistemic valence, or we rethink the idea of epistemic valence within a pragmatist framework. Those who've read the first two chapters of this dissertation should be able to guess that I take the third option.

It may help to consider an example of what I'm talking about, leading to clarifications in the next section. We might imagine two people who find themselves in the following Trolley Case:

Teamwork: Five people [are] on a track toward which an out-of-control trolley is heading. If it continues, it will kill them. However, if we redirect it away from them, it will go down another track where it will kill someone who is unable to be removed. In order to redirect the track, I must stand on a platform at the same time as my friend, because the platform is rigged up to a mechanism that turns the trolley and can only be triggered if we both stand on it. I have the intuition that we should stand on the platform, but my friend does not.

The case is designed to be similar to Trolley, discussed above, and hence for us to have the intuition that it is permissible to take the action. If the empiricist conception of intuitions is applied in this case, I am told to try to use some degree of non-rational persuasion to get my friend to agree with me and to take the action. But on the reasonable supposition that there is little time available to try to affect the friend's intuition in that way, there seems to be little in terms of method here. The rationalist may seem to be better off, insofar as the rationalist method of intuitions as data points can be operationalized through conversation and can be done by anyone who is willing to listen and participate. But these attitudes already presuppose that my friend is willing to listen and participate, and hence that my friend probably has some experience with moral philosophy. This seems unlikely, and it seems that therefore this is not a particularly

valuable method. The mere fact that calibration is possible, in Huemer's terms, should be worse than no consolation for us philosophers who see ourselves as involved in actual decision making.

Another reason to think that a Huemer-style solution is not plausible is the fact that it seems to result in the indeterminateness of a moral theory. Suppose that some answer to Teamwork is an intellectual seeming, and that Huemer or someone like him can say this. For example, suppose that Audi comes along, and has engaged in holistic reflection to determine that the friend should stand on the platform. Now further suppose that the rationalist were to try to convince the friend to comply. Is the method that the rationalist uses here a part of the rationalist moral theory, or is it not? It seems like the rationalist should count it as a part of the method of intuitions, but this means that any method that the rationalist can use to convince the friend could count as a part of the method of intuitions. But there may be an extreme diversity of methods that the rationalist can use to convince the friend. This means that if the rationalist says that the method of persuasion is a part of the rationalist method of intuitions, then the rationalist method is so diverse as to no longer provide something like a theory of the method. It is simply a narrow theory that tells us to reflect, and leaves the meaning and implications of that prescription radically open. It is, in other words, deeply indeterminate. But it seems that the rationalist's approach to the method of intuitions should be at least a little less indeterminate than this; it seems like the rationalist should be aiming at providing a final vocabulary for moral inquiry. This openness should be congenial to a pragmatist approach to the method of intuitions, but it seems hard to square with a broader rationalist approach.

My suggestion here, however, is that simply beginning the rationalist's method of intuitions, by working through a calibration of the particular intuitions that the friend has, is a way to acknowledge epistemic valence. The reality of the situation is that the friend will be

unlikely to be persuaded, but that our epistemic situation can be improved with the process of eliciting the friend's intuitions with the aim of working toward the resolution of the disagreement. It may seem odd to speak of merely trying to resolve a disagreement as an improvement of an epistemic situation, but this should seem less so if we keep in mind that even being able to elicit an intuition clearly is difficult, because intuitions are ephemeral and subject to framing effects. In an important sense, any eliciting of intuitions improves the epistemic situation of moral agents.

This way of thinking about the epistemic valence of intuition, as something that begins weak and through discourse becomes progressively stronger, steers a middle path between rationalism and empiricism, as these views have been discussed here. In a case like Teamwork, I would make everyone epistemically better-off if I try to persuade my friend, regardless of the results of doing so. Rationalists, crucially, cannot account for this because rationalists accept knowledge as the relevant standard for judging attempts at disagreements, and that is by hypothesis not satisfied in this approach to Teamwork. Empiricists, in contrast, have a hard time saying how this improves any epistemic situation, which I find to be implausible. In the next section, then, I develop this idea further into something more closely approaching a pragmatist account of intuitions and the method of intuitions.

3.5. Pragmatism and the natural kind theory of intuitions

The result of the previous two sections is that, for various methodological reasons, empiricist and rationalist conceptions of intuition cannot make sense of the actual methodologies that moral agents apply, and should apply, to real life cases of disagreements about moral intuitions, and cannot provide a more nuanced approach to epistemic valence in real situations. The reason for this is that empiricists and rationalists present extreme conceptions of the aim of

methodological moves in the face of disagreement. Empiricists about intuitions on one hand reject the idea that there could be a method because having a method for resolving disagreements requires that method's aiming at something stronger than emotional persuasion. Rationalists, on the other hand, cannot effectively operate the methods that they propose for resolving intuitional disagreements, in view of the overly demanding notion of intuitional success that is provided by the rationalist. For this conception of intuitions, intuitions are foundational and a priori, with their purported epistemic valences. Trying to provide a method for checking intuitions for these properties is simply too demanding in view of a large swath of demands that our situations place on us, even granting that this checking is possible.

Rationalists recommend attempts to reflect on our intuitions as a way to determine whether the intuition actually falls under the category of an intellectual seeming, or whether it is just a causally-valenced presentational state. Hence, the reason that this method does not help the rationalist is not because the method itself can play no role, but because the method cannot regularly be expected to measure up to the role that rationalism demands of it, in some of the most important cases. Similarly, the rationalist can extrapolate a reflective equilibrium style methodology from the intuition as data point view, and hence can provide a model for how we can trade off our intuitions to produce theories with fewer exception cases and a better internal fit. This method, however, is enormously intellectually-demanding, and requires systematic thinking that is not germane to the styles in which everyday moral reasoning occurs. But on the looser assumption that the aim of calibration is not merely to smooth over theories but to provide a tool for thinking about intuitions and their successive refinement over the course of discourse, with relatively little expectation of ending up with a theory at the end that can vanquish all challengers, this way of doing ethics could very well be something worth teaching people to use.

The key, it would seem, is to simply expect less of our method of intuitions; we might then find ourselves to be more satisfied with its relatively provisional and defective results.

The idea of intuition as simply one imperfect method among others for investigation of disagreements about ethical intuitions can help us to see how pragmatists can, contra the appearances suggested in the first section of this chapter, give an account of intuitions. We saw there that pragmatists have been generally quite skeptical of the idea of intuition because of its historical links with the ideas of the foundational and the a priori. But the approach to intuition that I pursue in this section shows us that the link can be severed. There is a question of whether we should accept a rationalist or an empiricist conception of intuitions, but even granted that that question is unanswered, we can say quite a lot about the role of the method of intuitions, and the relationship that that method has to real intuitional disagreements. In other words, even on the assumption that all of our intuitions are not epistemically-valenced, on the rationalist's idea of epistemic valence, there could be some role for our intuitions, if our intuitions aim not at the foundational or the a priori, but merely aim at being a back-up for understanding our actions and beliefs when no better answers are available. Hence, these methods are quite useful on the assumption that the broader epistemological project of intuitionism in metaethics is wrong-headed. If we want to make sense of our intuition practices, we ought to accept a pragmatist notion of intuitional success.

It is helpful to understand a bit of history to see how this is possible. I have criticized something that I called the empiricist conception of intuitions, but as it turns out, that is a bit of a misnomer. The Haidt-style approach to intuition in ethics is a kind of empiricism, but Haidt is not Quine, who is also associated with a broadly empiricist project. For Quine, empiricism is understood in a more classical sense, in which science and scientific investigation come as close

as anything to giving the final vocabulary of inquiry (recall the discussion of Quine's take on physics in Chapter 1, 1.3). Understood in this way, empiricism can be leveraged to provide an account of intuitions in ethics that is both more methodologically-respectable and in concert with a Rortian approach to ethics. Quine, cited in the first section of this chapter as highly skeptical about the idea of intuition in philosophy, was associated with naturalized epistemology, which views epistemology as a science of people's knowledge practices as a tool for making recommendations about better ways for us to pursue the sciences. This approach eschews normative epistemology and epistemologically-weighty notions such as the a priori or the foundational. Epistemology is viewed, here, not as providing a foundation for science, as many had conceived it in the past, but as being a tool for engaging with science (Quine 1969a, p. 78). Hence, on a Quinean view, the idea that an intuition could be an intellectual seeming that provided epistemic valence that was a sufficient condition for justifying a knowledge claim would be simply absurd. Yet the program of naturalized epistemology, as it developed past Quine, looped back toward intuitionism in the work of Hilary Kornblith (1996), who argued that a Quinean approach to epistemology could be compatible with a significant epistemic role for intuitions. For Kornblith, intuitions are rough-and-ready judgments about natural kind membership: a more-or-less well-trained judgment that some object is a member of a certain class (Kornblith 1996, pp. 134ff). Scientists, says Kornblith, often do this when confronted with a case that they are not sure about. We might imagine a geologist presented with a new kind of rock never before seen, and who forms a trained guess that the rock is igneous. When the geologist makes this prediction, they can then use this judgment to guide further inquiry. They may look first for properties of igneous rock, and may only look for properties of sedimentary rock when their judgment that the rock is igneous is ruled out.

Correspondingly, the natural kind approach to intuition in ethics would simply view intuitions in ethics as rough-and-ready judgments about whether a particular case in ethics is a member of, say, the class of morally permissible actions. Hence, upon considering some trolley case like Loop for the first time, we might describe what a person does by saying that they have gone about it by forming a more-or-less well-trained judgment that it is permissible to act. The person could then try to compare that judgment with other judgments about particular cases that they are more confident about. This might lead that person to make any number of conclusions. Three possible conclusions: that the intuition was right, that the intuition conflicts with some other, better intuition, that the intuition is close but that the actual answer is more nuanced and so the specific linguistic content of the claim accepted on the basis of the intuition would need to be revised.

One feature of the natural kind view of intuitions that is congenial to pragmatism is the way in which it situates intuition within the broader pragmatist approach to the scientific method. Pragmatists are often charged with being skeptics, insofar as they reject the idea that the scientific method provides a final method for investigating reality that answers to the very real demands of epistemic justification (Boghossian 2006, pp. 92-3). Here, the move is much like the earlier pragmatist move to reject the idea of final vocabularies, even while they must accept some vocabulary as final. We pursue science not because we believe that some day we will, through the rigid deployment of the scientific method, converge on the final description of the world, but because the scientific method has proved to be a useful way of trying to solve human problems. For Rorty,

there is no reason to praise scientists for being more “objective” or “logical” or “methodical” or “devoted to truth” than other people. But there is plenty of reason to praise the institutions that they have developed and within which they work, and to use these as models for the rest of culture” (Rorty 1991a, p. 39).

The value of scientific inquiry, on this view, is primarily in the institution of science, including the norms that govern the way that people go about studying natural phenomena. It is good if scientists follow the scientific method not because there is something intrinsically credible about that method, or because the method will converge on the truth if played out correctly, but because following the scientific method has proven to be relatively successful over time at helping communities and individuals to achieve their goals. In demythologizing science in this way, Rorty leaves open the possibility that much of what science does is not properly described as aiming at the truth. The value of science is in the virtues that the institution of science entails. Catherine Elgin (2017, pp. 127-9) agrees with Rorty and the pragmatist approach to science. Elgin expands these ideas to amount to the claim that normativity in science is moral, rather than epistemic: good science is science that is done by virtuous scientists who, while operating within epistemic communities, who act as though they are members of an epistemic kingdom of ends.

The connection between the natural kind conception of intuitions and the Rorty-Elgin conception of the scientific method is found in the fact that, on this conception, intuitions are better or worse to the extent that the method of appealing to natural kind intuition is called for in scientific investigation. Sometimes, scientists understand the subject matter that they are exploring quite well, and in such cases likely will make no appeal to intuition. Similarly, in moral cases, sometimes subjects need to make no appeal to intuition. It is not anything about the reliability of the method of intuition itself that allows this conclusion to be drawn, but simply our understanding of what sorts of methodological moves are possible and desirable, given the situational constraints on our inquiry and prior understanding of the subject matter, as we understand it. Hence, scientists rely on intuition more, and can do so with deeper confidence, when their subject matter is less well-understood. Similarly, in ethics.

This approach to intuitions can make sense of the deployment of the method of intuition as productive, given that we sometimes understand our subject matter poorly. As we come to understand our subject matter better, and feel more confident with several of the claims we make, our intuitions are either discarded or revised in favor of theories that serve the purposes of the intuitions more effectively. On such a view, conversely, it makes little sense to appeal to intuitions when some claim, or some subject matter, is already well-understood. This understanding of intuition, as I have already argued, comports well with the pragmatist's conception of science as one tool among others by which we come to understand phenomena that we are interested in. This understanding of intuition is also beneficial, in that it allows the pragmatist to make sense of the notion of epistemic valence in a way that leaves broader pragmatist epistemological critiques intact while still granting the epistemic valence of intuitions as having a robust, legitimate epistemic role in our lives. The natural kind conception of intuitions, in divorcing the epistemic authority of intuitions from the rationalist's foundational (in a strong, epistemological sense where a foundation is known through its status as intellectual seeming) and a priori aspirations can say why some intuitions are better than others. This is because some intuitions will more effectively stimulate and situate further scientific investigation, which is to say that, on the natural kind conception of intuitions, there will be cases where having an intuition about an action's being a member of a moral kind leaves us in a better epistemic position to investigate that moral kind, and its membership, than we would be if we had had some other intuitions.

Elgin's discussion is fruitful for the further reason that it explains how our practice of science, and by extension the practice of the method of intuition in ethics, can be valuable with no reference to strong philosophical properties like convergence on the truth. We might argue

that intuitions on the natural kind conception are best understood as felicitous falsehoods, or as judgments that we make that are epistemically justified because they are useful in guiding our investigations, and not because they are true, or because they have some other, similarly strong epistemic property (Elgin 2017, pp. 23ff). Consider the ideal gas law, $PV=nRT$, an equation that conceptualizes the relationships between various properties of gases. Elgin notes that the ideal gas law is false, in that actual gases do not always have properties that are predicted by the equation, and that even given that, the ideal gas law is a useful tool for estimating the properties of a specific quantity of gas, when all the values in the equation, except one, are known (Elgin 2017, p. 15). Elgin considers the ideal gas law to be a felicitous falsehood, because even though it is not even approximately true, and has a number of counterexamples, scientists regularly find the equation to be useful. Intuitions, under the natural kind conception of intuitions, are much like the ideal gas law in this regard. According to the methodological role that this conception grants to them, it does not matter, for the epistemic valence of an intuition, that the intuition is predictably false. Intuitions need not be true, and indeed need not care about the truth in any way, to have epistemic valence. We instead know the epistemic valence of intuitions retroactively, by looking at the history of some scientific investigation and seeing whether having that intuition was conducive to the understanding that has been progressively reached, or else know them inferentially, by making judgments about the potentially reliability of the intuitier who tokens the intuition. At no point in judging this is a claim about the truth conduciveness, foundational status, or a prioricity of the intuition; even on the supposition that there are true intuitions, whatever that means, we can get along just fine with only the class of the useful intuitions.

Before I move on to further elaborate the natural kind method of intuitions in a pragmatist framework, I want to address problems surrounding the term ‘natural kind’ in this context. In the cited Kornblith essay, the argument proceeds from two claims: first, that science is a matter of sorting physical objects into the natural kinds to which they belong, and second, that intuition as a practice is justified by its essential role in the first. From this arises a problem. Speaking of science as the investigation of natural kinds suggests a kind of realism about science which is unpalatable to pragmatists, because the language of natural kinds suggests that there is something natural, outside of the scientific discourse itself, to which that discourse responds. What I have argued so far is that pragmatists ought to cash out the idea of intuition and its epistemic role in virtue of Kornblith’s second claim. But what of the first, about the aim of science? Pragmatists want to accept the second claim, but accepting the first claim seems to be necessary to cash out the second claim, because only if science concerns itself with natural kind memberships would the method of intuition play its purported epistemic role. To name the problem according to language that Rorty criticizes, the idea of a natural kind might imply the idea that science responds to some metaphysically-independent notion of a world which is then categorized by scientific investigation. But, according to Rorty, that notion of a ‘world’ is nonsensical for pragmatists, insofar as some primordial conceptual ordering of these entities is a presupposition of our being able to scientifically investigate them in the first place (Rorty 1982c, pp. 14-15). What, then, could cash out the idea of a natural kind for a pragmatist? I think the pragmatist essentially has two options here. First, the pragmatist could try to establish a social constructionist conception of natural kinds, according to which natural kinds are fully metaphysically real, but metaphysically real because human practices have modified the world in such a way that the natural kinds are what they are. We make natural kinds become real by

talking about them and using them to understand the world, on this line. A second option is to decline to provide a theory of natural kinds that makes them metaphysically real, but that instead views natural kinds as a mere convenient way of talking about what we do in science. On this second option, natural kinds are not real in any metaphysical sense, but are theoretical posits that we use to guide scientific investigation (see Bird 2018, sec. 1.1.3; the claim is that the contrast between the first and second options here mirrors Bird's contrast between strong and weak conventionalism). I think that a social constructionist conception of natural kinds is possible, but I take the second option here. I think that pragmatists should treat natural kinds as practical posits that are justified, to the extent that they are justified at all, just insofar as talking about natural kinds is a useful tool for guiding scientific investigation (in other words, I think pragmatists should be weak conventionalists, in Bird's language). One can adopt this second option while remaining agnostic on whether there is a social constructionist account of natural kinds that is consistent with pragmatism. In other words, despite initial appearances, it is possible to account for the role of intuition as essential in investigating natural kind memberships (Kornblith's second claim) while rejecting the stronger claim about the role of kinds in science. The pragmatist can claim that science can investigate natural kind memberships even without endorsing any kind of realism about those kinds themselves. Rather, on this view, what science does is engage in processes of *kinding* observed phenomena, where the goal is to produce an ordering of those phenomena into kind-like groups that serve the practical purposes to which we put our science. Processes of *kinding*, then, it claims, make perfectly good sense even if there are no natural kinds. In contrast, a pragmatist who defends social constructionism about natural kinds would accept both of Kornblith's claims, and simply argue that the apparent difficulties with defending Kornblith's first claim can be overcome. Hopefully it is clear, then why it makes

sense to avoid social constructionist claims, from a pragmatist perspective: if we can get along just fine with processes of kinding that make no claim to the reality of natural kinds, then the further social constructionist claims would at best be pragmatically neutral, because even if the account worked, it would add nothing that was not captured in the processes of kinding that are more obviously the way to address the problem, on a Rortian approach to science.

Now to conclude this chapter. In further cashing out the idea of the natural kind conception of intuitions, and motivating it as a pragmatist conception, it will be helpful to discuss calibration, and to see why the natural kind conception does not fall to the calibration problem in the same way as the rationalist's conception does. To recall, the calibration problem holds that intuition must be verified against some non-intuitive standard in order to have some epistemic valence. We must be able to check our intuitions in order to show that they are reliable, insofar as some judgments that seem like intuitions, in the rationalist's sense, are false, or misleading. But the calibration problem is that checking our intuitions shows that the intuition itself is irrelevant, because whatever confirms the intuition, and not the intuition's epistemic status itself, is the thing that explains its epistemic valence. For instance, we might try to confirm an intuition by checking it against our theories. But if the theory confirms the intuition, then it is the theory itself that has epistemic valence. On the other hand, if the intuition is not calibrated, then it is not clear why it should have epistemic valence in the first place. So, for the rationalist, an explanation of the intellectual authority of an intuition is both necessary, and cannot be provided. The natural kind conception, however, entirely bypasses the calibration problem, while still making sense of the fact that we have epistemically-valuable practices of calibration. This is because the natural kind conception grants intuitions their epistemic role while simultaneously holding that intuitions aim at being tossed aside by being assimilated to our broader theories.

Natural kind intuitionists grasp gladly the claim that, when we confirm an intuition by checking it against our theories, we often show that the intuition is superfluous. But natural kind intuitionists deny that this fact shows that the intuition plays no epistemic role. The intuition played the epistemic role of guiding investigation into a case where the theory was applied, and can therefore show how the theory is stretched and expanded to cover the previously unknown case. There is no further expectation that, after inquiry has stopped, the intuition itself still occupies the same epistemic role it did when inquiry began. It does not matter whether further evidence is used to check the intuition, because the goal of providing intuitions in the first place is to effectively guide the search for better, non-intuitive evidence. Hence, the natural kind conception can make sense of the calibration methodology, while stopping short of shaping that method into a problem of any kind.

CHAPTER 4: A Pragmatist Metaethics? Reconstructing the Terrain

4.1. Introduction: plan of the chapter

Liberal ironism, which is a crucial tenet of Rorty's version of pragmatism, tells us to treat our own moral and metaethical commitments as incomplete, contingent, and radically amenable to rejection and revision. As such, it posits a kind of internal tension in the beliefs of a moral agent. Yet although it has these implications, pragmatism can also make sense of our actually having moral and metaethical beliefs. It is not so much that we cannot believe what we believe, if we are liberal ironists, but more that we must treat our beliefs with a bit of suspicion. This was seen as a potential problem in Chapter 1, but here, I will show that it is an important feature of the view, because it means that pragmatism can offer a more plausible account of the contents of moral discourse than can some standard metaethical views, which deny moral belief the space that it would seem it clearly occupies. Pragmatism can also account for the most plausible aspects of many of those practices. Pragmatism is hence shown as having the unique ability to absorb the best of the terrain, while rejecting the worst.

4.2. Non-cognitivism

Non-cognitivism is the metaethical view which says that moral sentences do not aim to state truths; in the jargon of the field, moral sentences are not truth-apt. Although the early non-cognitivists generally adopted this position because they saw moral sentences as semantically infirm when compared to sentences such as the sentences that constitute scientific theories (Ayer 1952, p. 107), others have gone to see this quality of moral sentences as independently plausible, given the poor prospects for showing that there are moral facts that could make a moral belief true or false (Blackburn 1984, pp. 188-9).

Non-cognitivism has been subject to persistent criticism because it seems to entail that our moral sentences do not function in our actual discourses as non-cognitivism says that they do. One way to state this is in terms of the embedding problem. If non-cognitivism is to be believed, and moral sentences are not truth-apt, then the following sentence is not truth-apt:

1. Stealing is wrong.

This sentence is instead read as an emotional expression toward the idea of stealing, or as a covert attempt to get other people to not steal things, or as the expression of the commitment to a norm against stealing. But any moral sentence will also have this property; this means that the following sentence is also not truth-apt:

2. Stealing the Mona Lisa is wrong.

This much is clear. But ordinary people would normally say that 2 could be inferred from 1. In virtue of knowing 1, it seems like I am now in a position to construct an argument that justifies concluding 2, because there would be a conditional relationship between 1 and 2, with 1 being the antecedent and 2 being the consequent. The argument might take the form of modus ponens, as follows:

1. Stealing is wrong.
3. If stealing is wrong, then stealing the Mona Lisa is wrong.
4. Therefore, stealing the Mona Lisa is wrong.

It seems like a metaethical account should not entail that this argument is invalid. But non-cognitivism is thought to entail exactly that. For note that neither the antecedent nor the consequent of the conditional premise are truth-apt. Yet nonetheless, the conditional itself seems to be truth-apt, and obviously so; it asserts an intuitively plausible inferential relationship between the two constituent sentences. But if the conditional is truth-apt, then the argument does not go through. The instance of ‘stealing is wrong’ that occurs in the first premise does not

clearly mean the same as the instance of ‘stealing is wrong’ that occurs in the second premise, because the second premise’s truth-aptness means that there is no neat linkage between the semantic properties of the conditional and the non-conditional premise. Hence, non-cognitivism entails that we cannot make this inference. But insofar as we are happy to grant that inference as a paradigmatic case of valid reasoning, that means that non-cognitivism has a serious flaw, and entails that people are generally wrong about how their moral sentences function in moral discourse (Geach 1965).

Of course, non-cognitivists have spent a lot of effort trying to answer these challenges, and have been successful to various degrees. In this section, I want to focus on Simon Blackburn’s (1984) quasi-realist version of non-cognitivism for reasons that will become apparent in a moment. Blackburn thinks that we can explain inferences like the inference from 1 and 3 to 4 by appeal to a logic of higher-order attitudes: the validity of the inference is explained by the fact that, contrary to appearances, 3 is actually not truth-apt, but instead is interpreted as a kind of expressive commitment to a relation obtaining between two sentences, each of which is itself not truth-apt. Inference between non-cognitive sentences is not analyzed as a logical notion, but as a practical notion (Blackburn 1984, pp. 193-5). Hence, the mere occurrence of the non-cognitive sentences in the conditional does not undermine the validity of the inference, because the semantics of the conditional have been revised so as to match the non-cognitive semantics of its constituent parts.

Perhaps a Blackburn-style strategy could be made to work, but I think we should resist it, and along with it the broader quasi-realist project. It is one thing to say that moral sentences and moral conditionals have an expressivist semantics that can be brought out through quasi-realist critique of moral discourse, but entirely another to posit a revision to the common-sense

rules of inference. I read a Rortian pragmatist as committed, at least as a first rule, to leaving common-sense attitudes toward our language intact. Hence, we should be strongly inclined to treat modus ponens in moral reasoning as the reasoning from assertion to assertion, rather than employing a higher-order attitudes approach to those inferences. This, of course, is not to say that this approach to moral semantics is incoherent; it is just to say that we have good pragmatic reasons to treat inferences along common-sense cognitivist lines. It also seems to me that the reasons for accepting non-cognitivist analyses in the first place, many of which appeal to some subjectivist critique of the possibility of moral truth values, are simply not strong enough to overturn this inclination. A subjectivism about moral sentences actually does not leave non-cognitivism about moral sentences as the only option, because a subjectivism about moral sentences is compatible with their having an extensional, Quine-Rorty semantics (again recall Chapter 1, 1.3) that allows moral modus ponens to go off without a hitch. Although this idea will get built out in later sections, I mean here something like a disquotational account of semantics, according to which, for any sentence p , p is true if and only if p (Rorty 1991c, pp. 137-8). On such an account, there is no complex question about the semantics of moral sentences as they are placed into conditionals; the truth values, insofar as they are specified disquotationally, are relaxed enough so that they are not relevant as to whether the inference goes off as valid. There is no problem with saying that (1) and (2) above entail (3), because the constituent pieces of all sentences have their meanings as a purely formal matter.

As an aside, I think the claim that Rorty-style pragmatism should accept the disquotational theory about truth needs some defense, because Diane Heney (2016, pp. 97-8), in a recent attempt to build out a pragmatist metaethics, argues that pragmatists should reject deflationism about truth and embrace social constructionism about truth. For Heney, we can't

abstract our truth-claims from our social practices, as a pragmatic matter, because our truth-claims just are social practices. What matters is whether there is a kind of responsibility taken for a claim if it is a truth-claim. Hence, insofar as the disquotational theory is merely formal and contains nothing normative, it robs truth-claims of what is most distinctive about them. I think that Heney is wrong about this, for two reasons. First, Heney offers no explanation of how this normative account of our sentences that is not disquotational can solve the embedding problem. But second, and more importantly, this seems to represent a misunderstanding of the disquotational theory. The disquotational theory does not aim to rob concepts of their normative nature, but aims to recast truth-claims in a formal scheme which is more germane to the level of analysis at which it is necessary. With all that Heney claims, there is no reason to think that a pragmatist cannot adopt something like a pluralist conception of semantics, treating truth-claims as social properties and as socially embedded here, and as purely formal theoretical constructs there.

The criticism of Heney shows that moral sentences may be governed by different semantics as the context varies. This is important for what follows. For instance, a moral sentence may turn out to have the form of a truth-apt assertion, yet is uttered primarily because of the non-cognitive force that the utterance would have. It's not hard to find examples where this seems to be the case. Imagine that a pro-choice partisan shouts 'a woman has a right to choose' at a pro-lifer holding a sign that says 'abortion is wrong'. Both of these state moral sentences, and both sentences would be read as cognitive and truth-apt if we simply left the surface form of the sentences stand, as I have suggested is both plausible given a pragmatist metaethics and a consequence of a good general attitude toward analyzing sentences. It also seems plausible to say that their utterances have something like a non-cognitive dimension. After

all, shouting ‘a woman has a right to choose’ at a political opponent could be well-understood as aimed at expressing a person’s rejection of the pro-lifer’s values; we should say the same for the pro-lifer’s sign. If anything, a surface-level account of the qualities of these utterances suggests that a relaxed account of moral semantics can live comfortably alongside the non-cognitive elements endemic to many examples of moral discourse.

Of course, I am not the first person to provide a so-called hybrid theory of moral semantics. Friends of non-cognitivism, like Michael Ridge (2006), have long posited such theories as a way to reconcile the non-cognitive elements of moral discourse with the semantic complexities that arise in those accounts.¹ Ridge, who defends a view that he dubs ecumenical expressivism, holds that moral sentences always express desires and sometimes express beliefs, and that, in moral sentences, the desires have logical priority over the beliefs (Ridge 2006, p. 309). Ridge cashes this out as the idea that moral sentences sometimes carry with them an anaphoric reference to a belief about moral fact which itself carries with it no assumption of the facticity of the belief (Ridge 2006, pp. 313-14). Although Ridge, in that paper, is hard to pin down on whether the Quine-Rorty view on semantics is the right one, that view seems to be most compatible with the usage that Ridge finds in the cognitive element in ecumenical expressivism. Although Ridge’s discussion provides a road map for the kind of semantic theory I think that a Rortian could endorse to avoid the muddle of quasi-realist semantics, I take issue with Ridge’s claim of priority of the non-cognitive element of moral sentences. Ridge holds this view because he wants to show that ecumenical accounts of moral semantics are incorrectly, in the literature, viewed as privileging the cognitive elements of moral discourse. If the cognitive element of a

¹ Although I focus on Ridge here, it is controversial whether Ridge counts as a proponent of a hybrid theory, or is just an expressivist. See Bar On et al. (2014) for some of the nuances here. In any case, Ridge’s work is hybrid enough for my tastes because it does try to provide the distinctive balance of a hybrid semantics.

moral sentence were always prior, then the cognitivist is left with the challenge of providing a non-reductionist explanation of how that sentence could refer to a moral property, a challenge which itself arises because of the connection between the cognitivist/non-cognitivist debate and the debate over moral properties arising out of Moore's open question argument. By prioritizing the non-cognitive element, Ridge's semantics allows reference without non-reductionism. Ridge can simply say that reference is explained by the non-cognitive elements of the moral language employed: a moral sentence like 'stealing is wrong' expresses disapproval of whatever property happens to be picked out by 'is wrong', and the cognitive element of the sentence is parasitic on that sentence's internal structure being disquotationally referred to by the non-cognitive element. If the cognitive element were prior, conversely, the non-cognitive element of approval would need to be derived from the surface-level assertion, but that notion is quite a lot less clear than the former (Ridge 2006, p. 306). Yet the idea of building out the non-cognitive element from the cognitive element, which Ridge thinks is implausible, is actually nothing like that. It is true that it can be hard to see how the non-cognitive element of the semantics could be derived from the cognitive element, just because the non-cognitive element would then seem to be subject to open question concerns, but actually, the idea of closing the question regarding motivation is made possible if we accept the broadly Rortian picture that has been hinted at in previous chapters. The error in Ridge's account arises from the fact that the argument against privileging the cognitive element assumes that the sentence, rather than the vocabulary, is the relevant unit of analysis. It is hard to see how a non-cognitive element could be derived from a cognitive element if all we have is a singular moral assertion, such as 'stealing is wrong', and hence hard to see how we could have an ecumenical cognitivism that speaks to the non-cognitivist's concerns. It is hard to see how something with non-cognitive motivational force is derived from a mere sentential

assertion, if all we can look at is a single sentence. Yet if moral sentences are semantically interconnected in holistic unities, as the Quine-Rorty view suggests, then inferring the non-cognitive element from ‘stealing is wrong’ could be a function of a broader commitment of endorsing various other sentences which are constitutively related to ‘stealing is wrong’. The wrongness of stealing might be, to take just a single, narrow example, constitutively related to the social disutility involved in stealing. Hence, a sentence like ‘stealing is wrong’ might be used to derive a non-cognitive association about stealing by placing the idea of stealing within a broader nexus of desires that the agent has, such as the desire to live in a stable society, or one that does not allow theft to disproportionately harm marginalized groups. Rather than requiring us to go in for an ecumenical cognitivism or an ecumenical expressivism, then, a holistic approach to moral semantics allows us to see how the cognitive and non-cognitive elements are negotiated in a delicate balancing of semantic relationships within the normative discourses that we actually engage in while stopping short of broad claims about which element is prior to the other.

The value of the holistic approach can be brought out by considering an objection. One might say that this picture of moral semantics cannot make sense of the motivational character of moral discourse. An ecumenical theory has to privilege the non-cognitive element because a theory that privileges the cognitive element would make the motivational element of ethical judgment extrinsic to the judgment (Ridge 2006; Blackburn 1998, p. 61). But this is unacceptable, for non-cognitivist reasons (Blackburn 1993, pp 168-70). But, so goes the objection, the holistic approach to ecumenical moral semantics I propose here also severs the link between moral judgment and motivation. If the Quine-Rorty response has the implication that motivations are merely contingently connected to moral reasoning, as breaking the proposed

tight link, then although that response might be good for other reasons, it cannot be a legitimate reconstruction of quasi-realism. Now of course we could bite the bullet, and say that we can just accept judgment externalism. Instead, however I want to argue that my proposal can accommodate the link between moral judgment and motivation, and hence serve as a reconstruction of some of quasi-realism's semantic ideas. Allen Gibbard, in discussing the link between motivation and moral judgment, appeals to the notion of internalization, rather than internalism, as Blackburn does in "How To Be An Ethical Anti-Realist" (1993). For Gibbard, the internal connection between judgment and motivation is a learned connection, which is formed by an adaptive tendency to develop understandings of norms that follows from broader biological imperatives which we human agents share with non-human animals (Gibbard 1990, pp. 68-71). But appeal to biological imperatives does not help as Gibbard thinks it does. Biological imperatives are potentially cognitive, or at least tied up with cognitive expressions. A particularly salient case here could be the case of thick ethical concepts, which are evaluative terms that have descriptive components. It might be argued that saying something like 'murder is cruel' states a thick ethical concept of cruelty (Väyrynen 2021). Cruelty is both non-cognitively valenced when used this way, and also has a clear descriptive component. But that descriptive component may itself be part of the adaptive tendency of the term; it affixes the non-cognitive valence to the ability to attach the non-cognitive valence to cases. This is not sufficient to refute quasi-realist theories. Blackburn rightly holds that externalists can account for some cases of moral judgment, and also that there are descriptive components to our ethical judgments. But Blackburn wrongly holds that this line is clear, as does Gibbard. After all, if appeal to biological imperatives is to support a non-cognitivist view, it must be clearly able to demarcate the non-cognitivist elements of internalization. Absent those, it leaves open a holistic semantics that

treats crucial elements of assertoric moral discourse as constitutively linked, in a web of moral ideas, with non-cognitivism's expressive components. The conclusion, then, is that the best account of non-cognitivist explanations of their supposed superiority on the question of moral motivation and its internal character support an approach to semantics that rejects both ecumenical expressivism and ecumenical non-cognitivism. Quasi-realism, along with its motivational and semantics claims, is better accounted for by a pragmatist approach to moral semantics, on the lines of the Quine-Rorty view.

4.3. Error theory

Error theory is a metaethical approach that attempts to debunk morality by showing that its conceptual aspirations far exceed its actual conceptual reach. More specifically, error theorists try to argue two claims: first, that morality is logically committed to some normatively-significant conception of moral objectivity, and second, that that conception of moral objectivity fails to obtain. These two claims are then thought to entail the conclusion that, conceptually-speaking, our moral practices are flawed. Richard Joyce (2001), probably the most-commonly read contemporary error theorist, cashes out the relevant conception of moral objectivity as a variety of motivational internalism. Joyce thinks that our concept of morality is non-negotiably committed to the reality of motivational internalism, which, recall, is the claim that, if a person has a moral belief, then that person also has a motive to act on that moral belief. His reasoning for this is that nothing else can explain the overridingness of morality that our practices are seemingly committed to. If a person had a moral belief that was seemingly not motivationally overriding, then they do not have the thing that is supposed to make moral belief ontologically distinctive. Joyce then argues that motivational internalism is false. Hence, there is an error buried deeply in our moral practices.

A common way of responding to error theories is to insist that the first claim is false, and that morality is not committed to the error theorist's notion of objectivity. Although one might plausibly believe this, such claims are not effective against error theories, because error theorists can merely insist that the proposed argument shows that there is some other notion of objectivity, but not a notion of objectivity that is properly moral. In other words, one cannot simply object to the error theorist's claim by pointing out that sometimes people seem to use moral terms in senses that are at odds with the error theorist's notion of objectivity. Nonetheless, it may temper the error theorist's enthusiasm for their claims to see that, even granting that moral discourse makes the purported claim to objectivity, there are options that are nearby to this purported claim that do not entail any error. The plausible view we end up with, on this approach to the problem, is that there could be a core notion of moral objectivity that entails error while at the same time lots of things that bear relevant similarities to moral objectivity that entail no such error.

Begin with Victor Moberger's clever discussion of John Mackie's (1977) founding statement of error theory. In his paper, "Not just errors: a new interpretation of Mackie's error theory," Moberger argues that there is textual support in Mackie to attribute to him a hybrid interpretation of moral discourse, where some moral claims are cognitive, and hence subject to the error theorist's critique, while other moral claims are non-cognitive, and hence not subject. For example, although Mackie thinks that moral sentences refer to objective properties which do not exist, he also thinks that some moral discourse is aimed at expressing people's allegiance to socially-mediated codes of behavior, as evinced by his using of moral language consistent with expressivism (Moberger 2017, p. 8). Moberger, however, does not explore these claims in any systematic way, beside suggesting that they render error theory more defensible.

Of course, Joyce can still insist that non-cognitive ‘moral’ claims are not properly moral, because morality just is what is intrinsically motivating. It may be the case that we can get along just fine with non-cognitive ‘moral’ claims because they serve our practical purposes, but those claims are something else, which we might call following convention schmorality. Schmorality is great for our purposes, but it is not morality, and so moves like the ones suggested by Moberger do not really answer the error theorist’s challenge. We simply would not understand a term that is schmoral as properly moral (Joyce 2001, p. 23ff). If someone were to assert ‘stealing is wrong’ yet not view themselves as having reason not to steal, then one would fail to grasp the concept of moral wrongness, although they may successfully grasp other concepts that are similar to moral wrongness.

So construed, I think there is no real way to refute the error theorist’s point. If one is really committed to understanding the concept of morality in this way, then it is hard to know what would convince them otherwise. I suggest, instead, we simply embrace the fact that the error theorist’s point must be granted. Instead, it is worth thinking about how error theorists need to think about integrating various normative conceptions that are nearby to the moral into a broader conception of the ethical life. If it turned out that moral discourse was in error, but that most of what is nearby to the moral but not actually moral was not in error because it had weaker conceptual aspirations, then that would show that error theory’s scope has to be traded off against its theoretical correctness. Either it is narrow enough to be true, yet does not cover the very important, closely related conceptions nearby the moral, or it is wide enough to include these things in its conception of the moral, yet false, because those nearby conceptions do not actually entail error.

Let me clarify what I mean by ‘nearby to the moral’. It may be helpful to distinguish morality, explicated in the narrow way in which Joyce explicates it, from normativity more broadly, which allows commitment to a norm which one does not see oneself as having a reason to abide by. Nearby notions to the moral are normative notions that are similar to morality, but that differ in their relative laxity. For instance, someone might use the concept ‘moral’ in the institutional, non-cognitive way that Moberger suggests. Others might use the concept ‘moral’ in the natural kind sense imagined in the coming discussion of Cornell realism. But the Cornell semantics doesn’t seem to entail motivational internalism, and the non-cognitivists accept internalism in a weaker sense that is explained by emotions, rather than by anything objective about the meaning of a moral concept. If we hold fast to the error theorist’s conception of morality, then this will be no trouble, and ruled out by the theory. But this means that the scope of error theory is so narrow as to include only a singular, relatively uninteresting notion of morality.

The import of this line of criticism can be expanded when we notice that, once the nearby senses of morality are being specified, the floodgates open to a diversity of nearby senses of morality. If we can observe normative notions that are similar to in some way, but not reducible to, the error theorist’s conception of the moral, then we expand the scope of what is nearby to the moral. This allows us to see our discourse as sometimes errant and sometimes not errant, depending on which terms we pick, given the broader purposes of normative discourse that we are a part of. Hence, when we make moral claims, we are necessarily in error, but when we make moral or nearby-moral claims, we are not necessarily in error. But as it turns out, actual moral discourse is more like the nearby-moral than the moral simpliciter. Hence, as I suggested above, the error theorist’s dilemma is between the scope of the theory and its correctness.

Notice, by the way, that pragmatism, which is more ecumenical and potentially more open to broader thinking about what sorts of things constitute the moral, can make a better sense of these claims than can error theory. After all, suppose that the pragmatist were to say that ethics properly deals with the nearby-moral, rather than merely the moral. They would be able to argue that the moral, as used by the error theorist, is a sub-part of the nearby-moral. Hence, the pragmatist can cleanly make sense of how some of our moral practices are constituted such that they entail errors. But they can also capture the more plausible thought that many of the things that we do, such as trying to convince people to be feminists while being indifferent to the truth of feminism, are not at all in error.

4.4. Cornell realism

It is clear, in the face of all of this, that pragmatism can be used to provide a more plausible conception of moral semantics than non-cognitivism and error theory, and hence that we have some initial reason to accept pragmatism. However, the conception of semantics that has been defended here, an extensional, disquotational semantics that is a respectable extension of what I called the Quine-Rorty view, is quite thin, and hence potentially uninteresting, even if true. After all, Quine (1979) argues that, even though this semantic view can be used to make sense of the semantic adequacy of moral terms, that nonetheless there is very little that can be positively said in ethics, because there is vast extensional diversity in how actual moral agents view the class of sentences that fall under moral concepts like ‘morally permissible’. Ethics is, in that way, importantly unlike science, and hence even granted the extensional semantics that has been provided here, very little of interest in moral reality would follow. It is the aim of this section to argue that Quine’s conclusions in that paper are overstated, and that even granted the

extensional diversity of moral concepts, it is possible to make interesting and convincing moral arguments that a wide variety of people will be able to understand and engage with.

The arguments for this claim are, I think, surprising, because they make reference to the so-called Cornell realist school of moral semantics. This is surprising for three reasons. First, this way of thinking about moral semantics analogizes moral concepts to natural kinds, and holds that we can investigate the memberships of moral natural kinds through a broadly scientific methodology. Yet a pragmatist, at least on the face of the matter, would be disposed to deny the idea that moral investigation is like scientific investigation. Second, the scientific methodology used to investigate moral natural kind membership is highly realist in its assumptions, insofar as it takes itself to be investigating moral properties that are viewed as natural properties of objects, which are had by those objects as an objective fact. Yet it seems that a pragmatist would be more inclined to treat moral properties, to the extent that they want to speak in those terms at all, as being social properties, or perhaps as projective, secondary properties of objects. Third, and more generally and worryingly, Cornell realism treats moral natural kinds on the Kripkean rigid designator model of natural kinds (c.f. Kripke 1980, p. 121), and hence on an essentialist model that Quine and Rorty would surely reject. On this model, the extension of a natural kind term would be locked in and hence constructed as an aspect of moral reality through the causal history of the term's being used in a certain way. Although this view surely allows the possibility of changing extensions through complex causal historical processes, it is on the face of it hard to see how rigidification as a process could be anything more than a way of simply picking out certain causal histories as contingently privileged, if the Kripke model is read through a pragmatist lens. After all, it is not as if pragmatists like Rorty will resist the idea that causal histories are a crucial determinant of the extension of a moral natural kind. It is only when that

idea is expanded to include some notion of essence, or linguistic privilege, that the pragmatist should jump ship. But Cornell realism seems to require that this ship not be jumped. So clearly the pragmatist has their work cut out for them, if they are to appeal to Cornell realism in a way that leaves crucial features of pragmatism intact.

I think the pragmatist's answer here is to attempt a reconstruction of Cornell realism that is parasitic on the loose conception of natural kinds that is presented in Chapter 3, 3.5. Recall that, in that section, I argued that pragmatism can fruitfully make sense of the idea of intuitions in ethics as aiming to sort ethical sentences into moral kinds. Having an intuition, on such a view, is a defeasible suggestion that some action or state of affairs falls under the extension of a moral concept. These suggestions are then explicitly dropped or modified when further investigation raises doubts as to the wisdom of construing the extension of the moral concept in that way. This way of thinking about moral natural kinds, then, requires nothing other than that we often, as a matter of our epistemic practice, try to place ethical sentences into various kinds, on the basis of our immediate judgments about what kinds those moral sentences fall under and then further investigation into whether that immediate judgment is best, on the pragmatist standard of 'best'. Therefore, if my contentions in Chapter 3 are to be believed, then there is a way forward for the Rorty-Quine view in tandem with the Cornell realist view. This is because the essentialism of the Cornell realist view is not actually essential to the Cornell realist view, and because stripped of that essentialism Cornell realism provides a plausible model for how our moral concepts can be sturdy enough to do practical work in ethics, even granted that the contents and extensions of those concepts are contingent, as they must ultimately be for the pragmatists.

To get more specific as to how this is possible, it is helpful to consider the results that Richard Boyd thinks arise from his defense of Cornell realism in “How To Be A Moral Realist” (1988). Boyd thinks that it is crucially important for ethics that ethical reasoning and scientific reasoning are very similar in their structures: both involve the investigation of metaphysically real natural kinds, and involve a kind of theory-laden intuitionist method of reflective equilibrium that can take us to knowledge about these kinds and the properties that are embodied. To take his example, Boyd thinks that he can demonstrate the truth of a version of consequentialism by arguing that human goods form a natural kind that is referred to by our concept of moral goodness; there is a homeostatic property cluster of things like health, wealth, happiness, etc., which are clustered together because we know certain more basic natural facts about what makes a human life go well, and we know these more basic natural facts in turn because we have done centuries of painstaking scientific investigation to figure out what makes human lives go well (Boyd 1988, pp. 203ff). Hence, when we use moral sentences that make reference to goodness, such as, for example, ‘Biden’s infrastructure plan is good’, we do so by referencing this homeostatic property cluster of the good things that we have picked out through our scientific investigation. I think that a pragmatist can agree with Boyd up to this point. This is the value of the pragmatist disquotational semantics that I have defended here so far: it can make ample good sense of substantive claims of moral goodness, insofar as substantive claims are trivially permitted by the disquotational truth schema, and anchored in reality through the causal historical properties of the idea of moral goodness that many people can see through just a basic understanding of human society. It also comports well with the holism of that view.

One issue, I think, with Boyd’s account is the notion of a homeostatic property cluster. For Boyd, the various human goods form a cluster property because there is a kind of natural

equilibration that occurs just because of human biological, psychological, and social nature that makes it the case that the natural kind of moral goodness includes them, and he thinks that we can know truths about these properties because we can observe human history to see that these goods actually do cluster in this way. To take the examples from the previous paragraph, health and wealth cluster together under the kind of moral goodness because a good human life must include both of these things in a kind of equilibrated balance; too much of one and the other is undermined. And note that health and wealth are merely two examples; the homeostatic property cluster that defines goodness, on Boyd's view, likely includes much more than these two elements. Yet I think the idea of homeostasis here is problematic, insofar as the contents of the properties picked out by goodness are themselves radically contingent. Boyd, for instance, thinks that there is a notion of health that can be seen by all (most?) as being an essential feature of goodness, but it's not clear that the contents of the idea of health actually are essential in any way. There are examples where notions of health and healthy behavior can be seen to diverge due to differing cultural features. Imagine a person who expresses joy regularly, with large smiles and lots of positive words toward friends and family members. In an American or European context, this person would be transparently viewed as expressing a healthy affect. Although we understand that people express emotions that do not accurately reflect how they feel, the external signs expressed by this person are associated with a healthy affect. However, in an East Asian context, expressing lots of joy would be viewed as an indication of an unhealthy affect; in those contexts, this would be associated with an unhealthy affect because East Asian societies prioritize the idea that expressing joy is socially harmful for those who are not feeling joyous (Chentsova-Dutton et al. 2014, pp. 345-6).² This means that the way in which people

² I am indebted to Jesse Prinz for introducing me to this literature.

would be disposed to apply the concept of health shows significant diversity and varies in regular ways across cultural barriers. Now surely Boyd has resources to explain examples like this: assuming this divergence is true, we discovered this through careful ethnographic study of mental health in cultural context, and that therefore, contrary to how it used to seem, health is promoted in a way that we could never have seen before. Hence, Boyd can appeal to the science/ethics analogy, and then claim that ethics often makes progress by better integrating scientific advance into its dictates. But this response may be less attractive than it seems at first. If scientific evidence is supposed to settle the constituent features of the property cluster of goodness, and the scientific evidence itself points toward irreducible diversity in that property cluster, we may wonder whether health is really an aspect of the property cluster of goodness. Boyd's analysis may have instead just shown that there are multiple candidate homeostatic property clusters each purporting to be goodness, each varying (at least) along the dimension of the causal history of the concept of health that they employ.

The combination of homeostasis with the internal diversity of the relevant cluster properties thus leaves Boyd with a dilemma. Either we accept his notion of homeostasis, or we reject it. If we accept it, then it turns out that we actually end up with multiple, overlapping causal histories of the kind moral goodness. That's unattractive for Boyd, because he aims to show that a fairly robust variety of moral realism is true, but this pluralism about the causal history of goodness would undercut the case for the extensional adequacy of moral goodness that is supposed to ground moral truths. In other words, moral causal regulation would no longer be significantly similar to scientific causal regulation; scientific kinds do not typically result in multiple candidate property clusters. Alternatively, Boyd could reject his notion of homeostasis, although rejecting this notion is tantamount to rejecting the naturalistic reduction that Boyd

thinks is distinctive about extensional definitions. Causal histories, without homeostasis, would not be unified enough to account for the possibility of scientific and moral realism. The best he could hope for, in that case, is disquotationalism and the Rorty-Quine view, and the extremely relaxed semantics that would follow from that.

Yet as it turns out, the Rorty-Quine view can readily incorporate the idea of causal regulation without ending up in the dilemma that Boyd does. That is because Rorty-Quine can grant the first horn of the dilemma, and accept homeostasis along with the causal chain pluralism that would arise from it. Recall that I said that accepting causal chain pluralism is problematic for Boyd, because Boyd takes himself to be arguing in favor of moral realism. Hence, it'd be a serious problem if homeostatic property clusters ended up with plural causal chains. But this is not so for the Rorty-Quine pragmatist semantics, because that view can treat the pluralistic element of the causal chains as simply a feature of the complex diversity of moral life. What the pragmatist can say is that there are multiple causal histories for terms like moral goodness, and that when we reason morally, we elide many of these differences for pragmatic purposes, and when eliding those differences leads to moral problems, we appeal to pragmatics to fix our sense of goodness. We, in other words, grant that the extensional semantics provided by Boyd rigidifies the extensions of the terms that we use, but that because the causal histories of the terms we use are so complex, we need to clarify which rigidification of the term we are using to ground discourse. Again, this is perfectly consistent with Boyd's approach to semantics, but because he ties his position to moral realism, the appeal to pragmatics to fix the context of the natural kind term would not be available to him.

One seeming problem that this approach to semantics can solve is that it can make sense of the logic of disagreement.³ On semantic pluralism, the logic of disagreement can seem to be a mystery. If diverging causal histories and their corresponding rigidifications are what lead the contents of moral kinds like goodness to become fixed, then individuals can transparently appear to be using the same terms when they disagree over whether, say, Biden's infrastructure plan is good, but actually be using the term good with different causal histories in mind. Further, if this is always a possibility, then it leaves people who use moral concepts like goodness in a poor epistemic position to know that they are really talking about the same thing; the depth of this problem is correspondingly proportionate to how divergent causal histories actually turn out to be. Although I do not provide a full answer to this problem, and would like to address it in a future paper, I think that Rorty gives us resources to respond to it. I can say that the pragmatics of situations allow, or rather require, people to elide differences in the causal histories of their concepts when they engage in discourse. As situations place more epistemically-weighty demands on people's semantic capacities, we have to work harder to take account of the pragmatics. If two people disagree over the extension and causal history of goodness with respect to just a few cases, then they may be quite close in terms of their semantics, and can transparently talk about the same thing. As the number or magnitude of the disagreements increases, their semantics become further apart, and they have to do some more significant clarification so that they can come to talk about the same thing. In his essay "Texts and lumps," Rorty writes that, in order to be talking about the same thing in some case, all discussants need is to be close enough to be able to articulate the differences in their ideas. They need not share the same concepts and semantic contents, but they need to be able to analyze each other's semantic

³ For putting this objection to me, I am indebted to Gary Ostertag.

contents and their concepts (Rorty 1991b, pp. 88-9). I think we should say something like this to the objection. If it is a problem that people's semantics do not overlap because that means that they do not share the concept in some important sense, the problem arises because of a tendency to analyze the semantics in an overly synchronic way. It seems likely, however, that disagreement is at least partly diachronic, in that it may take significant intellectual effort to even conceptualize the terms of disagreement. Although this answer entails that it is possible for people to really feel as though they are disagreeing but not actually be disagreeing, that seems to be no more of a problem than is the fact that people in general can be deeply confused whenever they attempt to disagree with each other, in a variety of ways. A thoroughly pragmatic approach to semantics can thus capture the possibility of disagreement even given divergent causal histories, and can say that it is possible to apply the predicate 'same concept' to those divergent causal histories. The error is in thinking that, in order to render the logic of disagreement intelligible, parties need to share the concept precisely right at the outset.

This shows that there is a significant difference between much of what is done in science and ethics, contra Boyd: we can't expect scientific kinds to remain plural like this for a significant period of time. Homeostasis is likely to be more robust in science than it is in ethics, because opposed conceptions of scientific kinds do not live together as comfortably as opposition between conceptions of moral kinds does. Indeed, it seems like scientific kinds are likely to converge more readily than moral kinds because scientific kinds are often created by eliminating the pluralism that it seems to be deeply hard to eliminate in moral kinds.

Recall that, at the beginning of this chapter, I pointed toward Quine's (1979) claim that the extensional diversity of moral concepts undermines the possibility of giving a genuine moral semantics. I promised to show that Quine was wrong about that. Here, we can see why. Quine

need not move from extensional diversity, which is genuine, to the impossibility of a real moral semantics, because the pragmatics of moral semantics, that is, the actual way in which people work on the way that they split moral notions into kinds in practice, specifically takes on board extensional diversity, and does so in a way that enables people to at least come to clarify, through discourse, what they are talking about. Indeed, it is quite ironic that Quine, who criticizes logical positivists in his work, famously in “Two Dogmas of Empiricism” (1951), ends up with a positivist skepticism about the possibility of moral discourse and semantics. A more thoroughgoing application of the Quinean semantics would prevent Quine from having to go to that implausible end point, and would also distinguish his unique body of work from that older view.

4.5. Constructivisms Humean and Kantian

So far, Chapters 3 and 4 of this dissertation have taken a long tour through the weeds of several theories in metaethics and moral epistemology, with the goal of showing that pragmatism has the resources to account for the reality of moral discourse, and the rationality of engaging in that discourse by appeal to intuitions. In doing this, it shows that pragmatism is not empty, pejoratively subjectivistic, and arational, as some have supposed. It shows, in other words, that pragmatists can provide a cogent metanormative framework in which moral discourse can occur. In this final section, then, one final task remains: to build up more substantively these insights into a satisfying account of how this metanormative framework can be leveraged to fill out a pragmatist conception of moral progress. After all, if we think of the natural kind theory of intuitions in ethics proposed in Chapter 3, what is argued there is that the method of intuition interpreted in the naturalized way that Kornblith interprets it can be given a full-throated pragmatist explanation. But it does not explain what it is for an intuition to be successfully

refined and reconstructed, on pragmatist lights. This section aims to explain that feature of intuitions by appeal to a theory of pragmatist constructivism. Although I have argued that pragmatists can adopt various aspects of Cornell realism, semantic deflationism, and intuitionism as part of their metanormative framework, I believe, and will argue, that if pragmatists want to adopt a more fully fleshed out normative framework, that they should be constructivists; in that sense, I agree with Robert Schwartz (2016) on this point. Schwartz argues that Dewey's writings on ethics can be interpreted as committing him to a broadly constructivist view of moral discourse. I show that, although Schwartz's paper is a great first step, that its conception of pragmatism should be expanded in light of my earlier discussion.

Before laying out the fundamentals of pragmatist constructivism and introducing the reader to Schwartz's view, however, I want to introduce some literature that provides something like a pragmatist account of moral progress, insofar as the aim of this section is to show that constructivism gives the pragmatists the best tools to think about progress. Two excellent philosophical works provide nearby conceptions that will help to fix ideas: Michele Moody-Adams's *Fieldwork in Familiar Places* (1997), and Phillip Kitcher's *The Ethical Project* (2011). Moody-Adams considers herself to be a moral realist and to be in the business of providing a robust notion of moral objectivity; however, the way in which she introduces what I will call a dialogical conception of moral progress positions her as a very strong ally, at minimum, with the pragmatists. For Moody-Adams (1997, pp. 169-177), moral discourse is a complex social function whereby moral agents air problems and search for solutions that would harmonize their various aims in a way that renders the projects of all to be satisfiable to the greatest possible extent. To engage in moral discourse, then, is to be a party to a social practice that has publicly-available rules for interpretation; although not writing specifically in reference to Moody-Adams,

we can view Elizabeth Anderson (2015, esp. p. 40) as joined in this project when she points out that the public-facing element of moral progress and moral interpretation must include the voices of those who are marginalized, and who are the beneficiaries of moral progress. From this perspective, then, moral progress is a kind of settling of practical problems through dialogical methods in a way that renders the problems of more people or groups to be satisfied in a way that is more satisfactory to them. Moody-Adams's conception of moral progress is functional, in that a moral change is progressive insofar as it renders social functions more easily and cheaply satisfied than other possible changes. Kitcher's conception of moral progress is also functional, although it specifies a very different notion of function than does the conception of Moody-Adams. For Kitcher (2011, pp. 218-229), moral progress is measured in terms of evolutionary progress: a moral change is progressive insofar as it renders evolutionary functions more easily and cheaply satisfied than other possible changes.

Rorty, however, who also styles himself a pragmatist, thinks that we have to accept a weaker notion of moral progress, which I will call the ethnocentric conception of moral progress. For Rorty, progress is always progress by our own lights; particular cultures, which are the location of a particular set of goals and aims, make progress insofar as they reach agreements where they satisfy their goals and aims more effectively, through whatever means or methods are most acceptable by their own lights (Rorty 1998b, pp. 172-3). Rorty's conception of moral progress is considerably thinner than that offered by Moody-Adams or Kitcher. For Kitcher, different cultural groups will tend to converge, to at least some extent, in what sort of aims and goals they take themselves to be required, insofar as the evolutionary needs of humans are relatively stable across those groups. Moody-Adams, by contrast, would say that the notion of discourse that explains how progress is possible is more substantive than would Rorty. Now, if it

turned out that the best option for accounting for moral progress were what Rorty offered, that would be a mark against the pragmatist, insofar as the pragmatist wants to provide a reconstruction of moral philosophy. Moral philosophy has to aim at something a bit more ambitious than that. For instance, Rorty gives no account, and indeed likely can give no account, of how the relevant cultural group is to be demarcated. Societies are made up of differing and overlapping sub-cultures, and if we are to make sense of moral progress, on a view like the one he provides, we need to come up with a way of individuating the relevant groups. But it's not clear how Rorty could do that. So it seems like there is motive to go toward an account like that of Moody-Adams or Kitcher. Yet the objection from the Rortian direction is that neither of these accounts can explain how moral progress is possible, given pragmatist assumptions. Actually, the argument might go, Moody-Adams and Kitcher are still stuck within a representationalist paradigm, whereby moral discourse attempts to represent some correct view about how evolutionary functions are best to be discharged, or about the contents of the dialogical norms that most effectively satisfy our aims. In this section, I leverage a detailed discussion of various kinds of constructivism to show that, contra this imagined Rortian reply, we could go in for some notion of moral progress that is more substantive than Rorty's ethnocentric conception, without taking on the more realist elements of the Moody-Adams and Kitcher conceptions.

The first point I want to argue is that, if they want it, pragmatists can have a Humean version of constructivism for free, or on the cheap, if they want it. I'll follow Sharon Street (2008) in conceiving of Humean constructivism as making two key claims. First, moral sentences are conative and connected with subjective desires, in a fully Humean conception of those ideas. Second, complex or derived moral sentences are "constitutively entailed" (Street 2008, p. 232) by other more basic moral sentences. Street's constructivism is holistic, insofar as

the lines between complex and more basic moral sentences need not be direct, and can be potentially confused. Further, Street's constructivism comports well with the semantics that has arisen in earlier sections of this chapter. Street, then, bites the bullet on objections that show that, according to this, it is possible to have a perfectly cogent racism, insofar as one's racist beliefs can be constitutively entailed by moral sentences that one accepts as more basic. Those more basic moral sentences cannot be bludgeoned away by moral argument, and insofar as all of these moral sentences are subjective, this is no problem for the theory. Street, of course, does think that it is possible for people's moral views to change over time, and that if this happens, it is because one simply comes to accept different constitutive entailment relations within one's moral views. So, when I said that a pragmatist can have Humean constructivism for free, the reason for that is clear: this way of thinking about moral construction requires nothing that is unacceptable in a pragmatist's metanormative framework. Causal histories, the deflationist moral semantics, and the natural kind conception of intuitions all explain how we might view certain moral beliefs as entailed by other moral beliefs that we already accept. Pragmatism, insofar as it attempts to treat the surface forms of our moral discourse as transparently as possible, can further explain how a moral sentence might be subjectively basic, for a person.

Further evidence that pragmatists can adopt Humean constructivism comes from the already-cited paper by Schwartz (2016, pp. 43-44), who argues that the Deweyan notion that, when engaged in moral discourse, we start with the beliefs that we already have and then try to both build out further beliefs from what we already have and subject the beliefs that we already have to some degree of critical scrutiny, can explain how pragmatists can make sense of the idea of objectivity in moral reasoning. We can read Schwartz's claim here as holding that moral discourse is the investigation of the constitutive entailment relations of the moral sentences that

we accept when we begin the process of moral discourse. Humean constructivism thus follows for pragmatists on the cheap.

Of course, Humean constructivism is a pretty minimal constructivism. The only standard on the values one may hold, on such a view, is internal consistency; this is the real meat of the constitutive entailment criterion. Other versions of constructivism are more ambitious, and offer the moral agent more than mere consistency among conatively constituted moral sentences as the progress criterion. I want to focus in particular on Kantian constructivism, likely the most popular alternative to Humean constructivism. The contrast between Humean and Kantian constructivisms mirrors the traditional philosophical divide between consequentialism and deontology, and hence seems ripe for pragmatist dissolution and reconstruction. Kantian constructivists like John Rawls (1980) and Christine Korsgaard (1996) agree with the Humean constructivists up to a point, specifically on the view that a person's or group's system of values is constituted through something like a procedure of constitutive entailment, also known in the literature as reflective equilibrium. However, Kantians diverge from Humeans on whether there is a thicker normative idea, itself necessarily a part of the construction procedure, that serves to impose a more stringent criterion on normative theorizing: Kantians say that there is such a thicker normative idea, while Humeans deny this. By far the most common thicker criterion for Kantian constructivists is the idea of the Kantian notion of the self as an autonomous self-legislator. Rawls, in his paper "Kantian Constructivism in Moral Theory" (1980), explains that the construction procedure is grounded by the idea of the moral agent as having the two moral powers of a sense of justice and a conception of the good. Persons, on Rawls's view, are conceived as being mutually disinterested in what is good for all others (a conception of the good), but as being mutually interested in establishing a fair scheme of social cooperation on

which all persons, to the maximal extent possible, can realize their own private conceptions of the good (a sense of justice) (Rawls 1980, pp. 525ff; Rawls 1985, pp. 233-4). Korsgaard expands on Rawls's metaphysics of the Kantian self and defends a more nuanced version of the argument; although the details don't necessarily matter that much here, it is perhaps worth quoting her at length to get the flavor for how the Kantian self can help to ground a conception of normativity:

Your humanity requires you to conform to some of your practical identities, and you can question this requirement as you do any other. Does it really matter whether we act as our humanity requires, whether we find some ways of identifying ourselves and stand by them? But in this case you have no option but to say yes. Since you are a human you *must* take something to be normative, that is, some conception of practical identity must be normative for you. If you had no normative conception of your identity, you could have no reasons for action, and because your consciousness is reflective, you could then not act at all. Since you cannot act without reasons and your humanity is the source of your reasons, you must value your own humanity if you are to act at all (Korsgaard 1996, p. 123).

Hence, the take-home for the Kantian constructivist is that we can avoid the more relativistic elements of Humean constructivism, such as the fact that it leaves open the possibility of a perfectly cogent racist, by appealing to a more metaphysically-substantive conception of who are the parties to the construction procedure. Try as they might, the perfectly cogent racist could not sustain that belief in moral discourse, given Kantian notions of persons as being constituted by the two moral powers.

If pragmatists can get Humean constructivism on the cheap, then they'd need to work quite a lot harder to explain how Kantian constructivism could be acceptable. Yet nonetheless, for reasons discussed in Chapter 2, pragmatists have a pragmatic reason to find a way to incorporate Kantian considerations into their moral theory, insofar as many people who are targeted by pragmatists for persuasion are themselves Kantian, and those people would be more likely to be persuaded if they could keep some feature of their Kantian views. Given the central

role that the Kantian self plays in Kantian constructivism, then, the question to pursue now is what sense can be made of the Kantian self within a pragmatist framework, given that the Kantian self as presented here seems to be just the sort of thing that Rorty would argue cannot be sustained given the commitment to avoiding philosophy in the pejorative sense discussed in Chapter 2.

Nonetheless, I think a pragmatist could give a place to a version of the Kantian self. Again consider what would be allowed by Street's method of constitutive entailment. This method states a requirement on moral beliefs that they either be basic for the agent in question or constitutively entailed by one of the agent's basic beliefs. Yet what exactly the basic beliefs are is left radically open; one might, on such a view, with no irrationality, be a committed racist, if one can get the constitutive entailment relations among one's other beliefs to match it in the right way. A plausible defense of Street then holds that it is simply extremely hard to get the constitutive entailment relations to work along with racist commitments in the right way, just because there is so much in the world that militates against having a cogent set of racist commitments. But if one can, in principle, construct a system of constitutive entailments that give place to racist beliefs, then surely we can also construct a system of constitutive entailments that gives place to the Kantian self. To do this is as if one takes a plausible set of constitutive entailments and adds an additional, further restriction onto it: that certain constitutive entailment relations be rejected in virtue of the fact that they undermine, to take just one example, the Kantian notion of autonomy. We might think of the story as follows: Humean constructivism sets a minimal constructivist floor for the moral beliefs that one might have, with that floor specified by constitutive entailment from one's basic moral beliefs, and the Rawlsian system of constructivism is then an optional add-on for those who take Rawlsian liberal values seriously.

Now obviously, Rawls and other Kantian constructivists will not want to think about deontology as an ‘optional add-on’ to a Humean model of constructivism. It would be said, to take just one possible line of criticism, that treating deontological values this way makes them heteronomous, and hence not really deontological. Imperatives that command categorically and which flow from the Kantian conception of the self, on the traditional way of thinking about these ideas, cannot be negotiable in this way. In stating these views, therefore, it is important to be clear about what exactly is implied by the idea of the Kantian constructivist’s ideas being cashed out in such a way. Although it is perfectly cogent for a person to reject the Kantian self, on these terms, for those who do not reject it, it is not exactly right to say that the Kantian self is negotiable. Moral agents, being historically conditioned, find themselves with a diverse amalgam of moral ideas, and for most of us, Kantian notions of the self are at least a partial feature of that. It is true that we cannot develop these points to produce truly categorical imperatives, in the way that a classical Kantian would want to, but it is also true that historicizing the Kantian self produces a valuable second-best option for the Kantian, if the choice is between the truly Kantian self, the historicized Kantian self, and the Kantian’s last place option of being fully Humean. In other words, a plausible reading of Kantian constructivism as aiming at the full notion of Kantian autonomy informally entails a pragmatist bargaining down to a weaker notion.

Such a bargaining down is not unprecedented for Kantian constructivists who take seriously criticism of the fuller, Rawlsian version of their position. I want to canvas two attempts to explain how Kantian constructivism could be rescued, given plausible critiques of the Kantian self: Sem de Maagt’s (2019) instrumentalist reconstruction of categoricity, and Onora O’Neill’s (1988) appeal to the circumstances of justice as grounding feature of Kantian constructivism. To prefigure the conclusion, I think that O’Neill’s answer is closer to the best

option than is de Maagt's, but as we will see O'Neill's project is overly ambitious, and must be toned down to produce a fully satisfying answer to these problems. In this way, the accounts work best when they are fit together.

Both writers note that the idea of universal categoricity is problematic, given familiar concerns about the historical contingency of the idea of the self, and more generally the presence of broad ethical pluralism about basic moral ideas. De Maagt's solution to this issue is to grant the Humean's assertion that all practical reasons are instrumental, but to argue that some among these instrumental reasons are categorical. The way this argument works is to point out that the categorical instrumental reasons are the reasons that are instrumental to the having of all other reasons. This is supposed to defuse the Kantian's insistence on the distinction between hypothetical reasons, which are reasons that we have given our desires that certain ends be brought about, and categorical reasons, which are reasons that we have regardless of our desires about which ends should be brought about. What de Maagt does is argue that there are reasons that we have regardless of which ends we desire, but that need appeal to nothing more substantive than the fact that all people have certain higher-order instrumental reasons regarding our first-order instrumental reasons, which are negotiable and totally hypothetical. Hence, the idea of a categorical instrumental reason (de Maagt 2019, pp. 288ff).

Although de Maagt's account is plausible as an explanation of why some reasons are deontic, it doesn't clearly explain the scope of those deontological reasons. Even granting that everyone has some higher-order instrumental reasons, it's not clear why that fact need show that there is enough overlap between the higher-order instrumental reasons that people have in order to make sense of a social contract. If, as seems plausible, some societies have people who hold different categorical instrumental reasons than the society from which de Maagt hails, then de

Maagt has an explanation only for how isolated individuals can be autonomous, in the sense entailed by the notion of the Kantian self. To explain how there could be a truly social construction, on Kantian constructivist lines, de Maagt needs to appeal to a supposed overlap of categorical instrumental reasons. But patently there is no such global thing; it is no more plausible to insist on this than it is to insist that, as Boyd does, goodness is a homeostatic property cluster. Hence, de Maagt could profit considerably by looking to O'Neill's Kantian constructivism, which provides a limited-yet-principled interpretation of overlapping categorical instrumental reasons. O'Neill achieves this by denying that the relevant construction class, or the group of individuals who are conceived as constructing moral principles, includes all persons, or all rational beings. O'Neill (1989, pp. 9-10) instead argues that the relevant construction class is not the individual and not the set of all individuals but is any set of individuals grouped together and unified by the particular circumstances of justice. The idea here is that people must construct principles for living together when their social and material circumstances place demands on them that force them to aim at designing such principles, under the auspices of what O'Neill calls an ethical plurality. In appealing to the particular circumstances of justice as a way to specify the construction class, O'Neill would provide the account that fleshes out de Maagt's account of categorical instrumental reasons by explaining how there could be sufficient overlap of categorical instrumental reasons without needing to stretch the idea further than it can go.

Despite this, problems remain for O'Neill's account. It might be said that the notion of the circumstances of justice are slipperier than O'Neill appreciates. Consider the fact that globalization has brought together previously far-separated corners of the globe under market forces. For instance, because of the way that global supply chains work, many of our products are made in China. This means that the people who produce those goods in China are a part of

our circumstances of justice, and we can plausibly owe things to each other. Fewer of my products, let's stipulate, are made in India, but enough that Indian manufacturing is still part of our circumstances of justice. But then suppose that the supply chains shift and the trade balance shifts toward India. There has been a kind of shift in the plurality: our plurality still contains both India and China, but now India is more central. So it seems that O'Neill's constructivism should be able to account for that. As a matter of fact, O'Neill can: she can grant that it is simply true that whomever could have claims against us is part of our plurality, and we simply work that out by applying something like an idea of the transitivity of plurality membership. But now it is plausible to worry that this means that the plurality is really just all of humanity. China engages in trade with North Korea, and hence North Korea and China would seem to be part of the same plurality. But it also seems like plurality membership should be transitive, because it is inevitable that, as a part of the microcausal economic process, American companies interact with Chinese companies that do business with North Korea. So this means that any person who does business with India or China is part of the American plurality. But now note even further that this is true for any country that does business with America. So it seems like we have discovered that we are all part of the same 'plurality': humanity. The idea seems to collapse, and we lost the aspect that seemed to make O'Neill's constructivism unique. O'Neill could argue that, within all of humanity, as plurality, we'd still get the right moral conclusions that, say, slavery and trafficking are wrong, because persons within a plurality cannot be reasonably expected to consent to their own domination; yet even given this, because this response endorses the idea of minimal rationality, it seems that proposing the idea of minimal rationality as an idea that unites all of humanity would run into the exact same concerns about universalizing that motivated the Rawlsian constructivist attempt to bypass Kantian problems in the first place. In other words,

without a more particular and exclusive notion of ‘plurality’, O’Neill cannot make sense of how Kantian constructivism provides a broadly plausible ethical ideal.

It seems that O’Neill needs to endorse a stricter notion of plurality membership, and hence a narrower view of what the relevant construction class is. Without providing an answer to this question that will satisfy all, it pays to note that this is a point at which it would be helpful for the Kantian constructivist to think really carefully through the fact that they get Humean constructivism for free, and that their arguments amount to arguments for the adding-on to that theory. They can avoid this dilemma by simply granting that the Kantian self is historically conditioned, and so that it is much more ethnocentric than it has previously seemed that it must be. Perhaps the Kantian should content themselves, then, with the fact that they can come up with quite convincing arguments that, within certain narrowly-defined contexts, the Kantian self, and hence Kantian constructivism, is a powerful live option, and that de Maagt’s notion of categorical instrumentality explains how certain aspects of our autonomy can be non-negotiable, again within context. But this notion, it follows, is perfectly compatible with a notion of pragmatism that takes on board Humean constructivism’s commitments, as I have suggested that it has good reason to. It turns out, then, that pragmatists can endorse a wide variety of constructivisms, and that those varieties turn out to have a common basis.

These claims can be brought into relief by considering Sabina Lovibond’s criticism of Rorty’s approach to ethics as overly ethnocentric, insofar as it requires that the criticism of social practices be too narrowly focused on the tightly-delineated norms of the *ethnos* itself, to the detriment of attention to the ways in which broader, even global norms condition those tightly-delineated norms. Lovibond argues that feminist criticism of patriarchal social practices, for example, cannot be properly conducted by careful attention to the tightly-delineated norms of

some particular society, because groups relate to each other internationally in potentially ethically-significant ways. Lovibond writes:

Consider for example such mind-boggling, yet urgently necessary undertakings as the global redistribution of wealth and resources, the reallocation of work and leisure, the prevention of war and environmental destruction. Well, no doubt we shall be told that there is something *passé* in the very habit of mind which can still frame this kind of classically humanist agenda, given the alleged ‘exhaustion’ of all our political traditions (MacIntyre) and the extinction of any shared ‘nostalgia for the unattainable’ (Lyotard). But on the other hand, if there can be no systematic political approach to questions of wealth, power and labour, how can there be any effective challenge to a social order which distributes its benefits and burdens in a systematically unequal way between the sexes? (Lovibond 2015a, p. 22)

Lovibond may well be right that criticism of injustice may require us to move beyond our *ethnos* in a way that a Rortian may find uncomfortable. I do think, however, that it is implausible to suggest that a Rortian should flee from this discomfort. A pragmatist approach to discourse and moral progress tells the pragmatist to try to extend their rhetorical framework to whomever could be convinced to accept it, and when nations come into relationship with each other through processes of globalization, it is not implausible to say that pragmatists have reason to try to extend their framework in those contexts. As the normative significance of national boundaries and the world becomes flatter, it becomes more and more reasonable to expect people outside of one’s *ethnos* to care about what goes on in there, and vice versa. To deny that is to require an implausibly narrow approach to norm-construction, and to land in exactly the problem that O’Neill lands in. Hence, although Lovibond’s criticism is initially plausible, ultimately it provides fuel for the pragmatist’s critique, insofar as it forces us to see that pragmatism, as Rorty imagines it, actually demands that we treat people outside of our *ethnos* with at least some degree of consideration, and conversely that their practices be taken up as objects of criticism. Whether they accept is, of course, not something that a pragmatist, nor anybody else, whatever their metanormative commitments, can guarantee.

Now that the framework for a pragmatist version of constructivism is developed, we can say something about moral progress, which was the stated purpose of this discussion. Moody-Adams and Kitcher are pragmatist-adjacent philosophers who have proposed various functionalist conceptions of moral progress: for Moody-Adams the functions in question are dialogical, and for Kitcher, the functions in question are evolutionary. I wondered, however, whether these notions could avoid falling into representationalism, because they require that the functions in question are stable enough to explicate prior to the debate in question. Conversely, we wanted to avoid falling into an ethnocentric account of progress, because that account is not ambitious enough to count as a legitimate reconstruction of the aims of moral philosophy. Whatever else is true, we can't have a conception of moral progress that makes no sense of the reasons that we may have to expand our *ethnos*. The account of constructivism that I have outlined here, then, is more effectively pragmatist than the functionalist one that has proved attractive to Moody-Adams and Kitcher, insofar as it can accommodate an anti-representationalist notion of moral function that does not appeal to dialogical success or to evolutionary function to be cashed out, yet is more substantive than the account offered by the Rortian ethnocentrist. It can make very good sense, on pragmatist lines, why our moral discourse often aims at those outside of our own *ethnos*, while still being beholden to a pragmatist notion of thinking of moral discourse as an ongoing process that we find ourselves as a part of at specific historical positions. On my view, to make moral progress is to find solutions to moral problems, whether that is cashed out in terms of evolutionary or dialogical success, or something else entirely, while appealing to shared moral understandings that could provide the basis of agreement. One crucial element of many people's moral understandings, however, turns out to be the notion of the Kantian self. My version of constructivism thus endorses the more ambitious

notions of the Kantian constructivist when doing so would not alienate individuals from that ongoing process of developing shared moral understandings, yet tempers them in view of the fact that those ideas have, for many people, a deeply checkered history.

CONCLUSION: Pragmatism, Reconstruction, and Philosophy on the Cusp of the Real World

In a sense, the results of this dissertation are modest: it demonstrates little more than that a pragmatist can, contra Rorty's views expressed in *Contingency, Irony, and Solidarity*, do moral theory and metaethics in a way that allows them to make substantive contributions to debates about topics that would be traditionally be ignored by pragmatists. In other words, Rorty's (1991d) distinction between democracy and philosophy in the pejorative sense is a contextual vestige that can be done away with should the pragmatist choose to orient themselves in a way that makes that distinction irrelevant. But what the dissertation does not show is that anything like a 'pragmatist normative ethics' or a 'pragmatist metaethics' follows; while a pragmatist can take a position coherently and contribute to debates in normative ethics and metaethics, pragmatists are not compelled, by virtue of the basics of their position, to accept any particular view. As it turns out, though, this is just fine, because while pragmatists cannot claim that some moral theory is the pragmatist moral theory, they don't need anything more than reconstruction to do what is needed of a moral theory. A moral theory has multiple aims, as we saw in Chapter 1, but one of those aims is to develop methods for settling debates that can win at least some agreement. And the method of reconstruction demonstrated throughout can do this even better than can a more traditional approach to moral theory. While a full argument for this conclusion would require the development of a fifth chapter, rather than a conclusion, the idea is that pragmatism and the idea of reconstruction is better placed to develop normative approaches to practical problems that avoid the dogmatism of the more traditional approach that attempts to hold some moral theory as being the right one. Take the example of medical ethics, where principles like beneficence and autonomy are regularly appealed to for the purpose of prescribing a doctor's action in some case. A pragmatist, reconstructionist approach to moral theory better

explains how a principle of, say, autonomy could be weighted off against some other principle, in virtue of the fact that pragmatism offers a highly plausible framework for thinking about deontological ideas like autonomy in a way that can institutionalize and justify a narrow procedure for doling out exceptions to the principle. In other words, it is because pragmatism can reconstruct moderate deontology within a Rawlsian framework, on the lines of the account developed in Chapter 2, that it provides a better approach to medical ethics than does a purer commitment to moderate deontology. A Rorty-style approach to moral theory, counterintuitively, can provide a framework for the resolution of clinical dilemmas for something approaching a principled reason: it treats moral philosophy as a tool for the resolution of problems, as demanded by the real world, and avoids engaging in theorizing that requires more substantive commitments that would be required by the former. Pragmatist reconstruction brings philosophy just a bit closer to the cusp of the real world.

BIBLIOGRAPHY

- Alexander L. 2000. Deontology at the threshold. *San Diego law review*. 37(4):893-912.
- Alexander L, Moore MS. 2020. Deontological ethics. In: Zalta EN, editor. *Stanford encyclopedia of philosophy*. Winter 2020 edition. <<https://plato.stanford.edu/entries/ethics-deontological/>>.
- Anderson E. 2015. Moral bias and corrective practices: a pragmatist perspective. *Proceedings and addresses of the American Philosophical Association*. 89:21-47.
- Audi R. 2004. *The good in the right: a theory of intuition and intrinsic value*. Princeton (NJ): Princeton University Press.
- Ayer AJ. 1952. *Language, truth, & logic*. New York (NY): Dover. First published: 1936. London (UK): Victor Gollancz Ltd.
- Bar-On D, Chrisman M, Sias J. 2014. (How) is ethical neo-expressivism a hybrid view? In: Fletcher G, Ridge M, editors. *Having it both ways: hybrid theories and modern metaethics*. Oxford (UK): Oxford University Press. 223-247.
- Barnes E. 2016. *The minority body: a theory of disability*. Oxford (UK): Oxford University Press.
- Bealer G. 1998. Intuition and the autonomy of philosophy. In: DePaul M, Ramsey W, editors. *Rethinking intuition: the psychology of intuition and its role in philosophical inquiry*. Lanham (MD): Rowman & Littlefield. 201-240.
- Bird A. 2018. Natural kinds. In: Zalta EN, editor. *Stanford encyclopedia of philosophy*. Spring 2018 edition. <<https://plato.stanford.edu/archives/spr2018/entries/natural-kinds/>>.
- Blackburn S. 1984. *Spreading the word: groundings in the philosophy of language*. Oxford (UK): Oxford University Press.
- Blackburn S. 1993. How to be an ethical anti-realist. In: *Essays in quasi-realism*. Oxford (UK): Oxford University Press. 166-181. First printed in: *Midwest studies in philosophy*. 12(1): 361-375.
- Blackburn S. 1998. *Ruling passions: a theory of practical reasoning*. Oxford (UK): Oxford University Press.
- Boghossian P. 2006. *Fear of knowledge: against relativism and constructivism*. Oxford (UK): Oxford University Press.
- Boyd R. 1988. How to be a moral realist. In: Sayre-McCord G, editor. *Essays on moral realism*. Ithaca (NY): Cornell University Press. 181-228.

- Brennan S. 2009. Moderate deontology and moral gaps. *Philosophical perspectives*. 33(1):23-43.
- Cappelen H, Plunkett D. 2020. Introduction: a guided tour of conceptual engineering and conceptual ethics. In: Burgess A, Cappelen H, Plunkett D, editors. *Conceptual engineering and conceptual ethics*. Oxford (UK): Oxford University Press. 1-26.
- Chentsova-Dutton YE, Ryder AG, Tsai J. 2014. Understanding depression across cultural contexts. In: Gotlib IH, Hammen CL, editors. *Handbook of depression: third edition*. New York (NY): Guilford Press. 337-354.
- Cummins R. 1998. Reflection on reflective equilibrium. In: DePaul M, Ramsey W, editors. *Rethinking intuition: the psychology of intuition and its role in philosophical inquiry*. Lanham (MD): Rowman & Littlefield. 113-127.
- Dancy J. 2014. Intuition and emotion. *Ethics*. 124(4): 787-812.
- Daniels N. 1979. Wide reflective equilibrium and theory acceptance in ethics. *Journal of philosophy*. 76(5):256-282.
- Davidson D. 1967. Truth and meaning. *Synthese*. 17(3):304-323. Reprinted in: 2006. *The essential Davidson*. Oxford (UK): Oxford University Press. 155-170.
- de Lazari-Radek K, Singer P. 2010. Secrecy and consequentialism: a defense of esoteric morality. *Ratio, new series*. 23(1):34-58.
- de Maagt S. 2019. Why Humean constructivists should become Kantian constructivists. *Philosophical explorations*. 22(3):280-293.
- Dewey J. 1920. *Reconstruction in philosophy*. New York (NY): Henry Holt and Co.
- Dewey J. 2018. *Democracy and Education*. Gorham (ME): Myers Education Press. First edition: 1916. New York (NY): Macmillan.
- Eggleston B. 2013. Rejecting the publicity condition: the inevitability of esoteric morality. *The philosophical quarterly*. 62(250):29-57.
- Elgin CZ. 2017. *True enough*. Cambridge (MA): The MIT Press.
- Ellis A. 1992. Deontology, incommensurability, and the arbitrary. *Philosophy and phenomenological research*. 54(2):855-875.
- Frances B, Matheson J. 2019. In: Zalta EN, editor. *Stanford encyclopedia of philosophy*. Winter 2019 edition. <<https://plato.stanford.edu/entries/disagreement/>>.

- Fraser N. 1991. From irony to prophesy to politics: a response to Richard Rorty. *Michigan quarterly review*. 30(2):259-266.
- Fricker M. 2013. Styles of moral relativism: a critical family tree. In: Crisp R, editor. *Oxford handbook of the history of ethics*. Oxford (UK): Oxford University Press. 793-817.
- Geach PT. 1965. Assertion. *Philosophical review*. 74(4):449-465.
- Gibbard A. 1990. *Wise choices, apt feelings: a theory of normative judgment*. Cambridge (MA): Harvard University Press.
- Greene J. 2014. Beyond point-and-shoot-morality: why cognitive (neuro)science matters for ethics. *Ethics*. 124(4): 695-724.
- Haidt J. 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review*. 108(4): 814-834.
- Hare RM. 1981. *Moral thinking: its levels, method, and point*. Oxford (UK): Oxford University Press.
- Haslanger S. 2005. What are we talking about? The semantics and politics of social kinds. *Hypatia*. 20(4):10-26.
- Heney DB. 2016. *Toward a pragmatist metaethics*. New York (NY): Routledge.
- Huemer M. 2005. *Ethical intuitionism*. London (UK): Palgrave.
- Huemer M. 2009. A paradox for weak deontology. *Utilitas*. 21(4):464-477.
- Hyde D, Raffman D. 2018. Sorites paradox. In: Zalta EN, editor. *Stanford encyclopedia of philosophy*. Summer 2018 edition. <<https://plato.stanford.edu/archives/spr2018/entries/sorites-paradox/>>.
- James W. 1891. The moral philosopher and the moral life. *International journal of ethics*. 1(3): 330-354.
- Joyce R. 2001. *The myth of morality*. Cambridge (UK): Cambridge University Press.
- Juzaszek M. 2014. Normativity of moral intuitions in the social intuitionist model. In: Brand C, editor. *Dual-process models of moral psychology: interdisciplinary approaches to theoretical, empirical, and practical considerations*. Wiesbaden (DE): Springer. 57-76.
- Kagan S. 2001. Thinking about cases. *Social philosophy and policy*. 18(2):44-63.
- Kamm FM. 2007. *Intricate ethics: rights, responsibilities, and permissible harm*. Oxford (UK): Oxford University Press.

- Kant I. 1998. Critique of pure reason. Guyer P, Wood AW, translators and editors. Cambridge (UK): Cambridge University Press. First printed in: 1781.
- Kitcher P. 2011. The ethical project. Cambridge (MA): Harvard University Press.
- Kornblith H. 1998. The role of intuition in philosophical inquiry: an account with no unnatural ingredients. In: DePaul M, Ramsey W, editors. Rethinking intuition: the psychology of intuition and its role in philosophical inquiry. Lanham (MD): Rowman & Littlefield. 129-141.
- Korsgaard CM. 1989. Personal identity and the unity of agency: a Kantian response to Parfit. *Philosophy & public affairs*. 18(2):101-132.
- Korsgaard CM. 1996. The sources of normativity. Cambridge (UK): Cambridge University Press.
- Kripke SA. 1980. Naming and necessity. Cambridge (MA): Harvard University Press. First printed in: 1972. Davidson D, Harman G, editors. *Semantics of natural language*. Dordrecht (NL): D. Reidel Publishing. 253-355.
- Kukla Q. 2015. Medicalization, 'normal function,' and the definition of health. In: Arras JD, Fenton E, Kukla Q, editors. *The routledge companion to bioethics*. New York (NY): Routledge. 515-530.
- Larmore C. 2003. Public reason. In: Freeman S, editor. *The Cambridge companion to Rawls*. Cambridge (UK): Cambridge University Press. 368-393.
- Liao SM. 2007. A defense of intuitions. *Philosophical studies*. 104(2):247-262.
- Liao SM. Manuscript. Threshold deontology: lessons from vagueness. 1-37.
- Lovibond S. 2015a. Feminism and postmodernism. In: *Essays on ethics and feminism*. Oxford (UK): Oxford University Press. 18-44. First printed in: 1989. *New left review*. I/172:5-28.
- Lovibond S. 2015b. Feminism and pragmatism: a reply to Richard Rorty. In: *Essays on ethics and feminism*. Oxford (UK): Oxford University Press. 45-65. First printed in: 1992. *New left review*. I/193:56-74.
- Mackie JL. 1977. *Ethics: inventing right and wrong*. London (UK): Penguin.
- MacLeod CM. 2014. Applying justice as fairness to institutions. In: Mandle J, Reidy DA, editors. *A companion to Rawls*. Hoboken (NJ): John Wiley & Sons. 164-184.
- Manne K. 2016. *Down girl: the logic of misogyny*. Oxford (UK): Oxford University Press.

- Moberger V. 2017. Not just errors: a new interpretation of Mackie's error theory. *Journal for the history of analytical philosophy*. 5(3):1-12.
- Moody-Adams MM. 1997. *Fieldwork in familiar places: morality, culture, and philosophy*. Cambridge (MA): Harvard University Press.
- O'Neill O. 1988. Constructivisms in ethics. *Proceedings of the Aristotelian society*. 89:1-17. Reprinted in: 1989. *Constructions of reason: explorations in Kant's practical philosophy*. Cambridge (UK): Cambridge University Press. 206-218.
- Okin SM. 1991. *Justice, gender, and the family*. New York (NY): Basic Books.
- Parfit D. 1984. *Reasons and persons*. Oxford (UK): Oxford University Press.
- Parfit D. 2011a. *On what matters, volume one*. Oxford (UK): Oxford University Press
- Parfit D. 2011b. *On what matters, volume two*. Oxford (UK): Oxford University Press.
- Paul LA. 2014. *Transformative experience*. Oxford (UK): Oxford University Press.
- Peirce CS. 1992. Questions concerning certain faculties claimed for man. In: Houser N, Kloesel C, editors. *The essential Peirce: Selected philosophical writings, Volume 1 (1867-1893)*. Bloomington (IN): Indiana University Press. First printed in: 1868. *Journal of speculative philosophy*. 2(2):103-114.
- Piper AMS. 1978. Utility, publicity, and manipulation. *Ethics*. 88(3):189-206.
- Pust J. 2000. *Intuitions as evidence*. New York (NY): Routledge.
- Quine WV. 1951. Two dogmas of empiricism. *Philosophical review*. 60(1):20-43.
- Quine WV. 1960. *Word and object*. Cambridge (MA): MIT Press.
- Quine WV. 1961. Notes on the theory of reference. In: *From a logical point of view: second revised edition*. Cambridge (MA): Harvard University Press. 130-138. First edition: 1953. Cambridge (MA): Harvard University Press.
- Quine WV. 1969a. Epistemology naturalized. In: *Ontological relativity and other essays*. New York (NY): Columbia University Press. 69-90.
- Quine WV. 1969b. Natural kinds. In: *Ontological relativity and other essays*. New York (NY): Columbia University Press. 114-138
- Quine, WV. 1979. On the nature of moral values. *Critical inquiry*. 5(3):471-480.

- Ramberg B. 2009. Richard Rorty. In: Zalta EN, editor. Stanford Encyclopedia of Philosophy. Spring 2009 edition. <<https://plato.stanford.edu/entries/rorty/>>.
- Rawls J. 1980. Kantian constructivism in moral theory. *Journal of philosophy*. 77(9):515-572.
- Rawls J. 1985. Justice as fairness: political not metaphysical. *Philosophy & public affairs*. 14(3):223-251.
- Rawls J. 1999. *A theory of justice: revised edition*. Cambridge (MA): Harvard University Press. First edition: 1971. Cambridge (MA): Harvard University Press.
- Rawls J. 2005. *Political liberalism: expanded edition*. New York (NY): Columbia University Press. First edition: 1993. New York (NY): Columbia University Press.
- Ridge M. 2006. Ecumenical expressivism: finessing Frege. *Ethics*. 116(2):302-336.
- Rorty R. 1979. *Philosophy and the mirror of nature*. Princeton (NJ): Princeton University Press.
- Rorty R. 1982a. Introduction: pragmatism and philosophy. In: *Consequences of pragmatism*. Minneapolis (MN): University of Minnesota Press. xiii-xlvii.
- Rorty R. 1982b. Pragmatism, relativism, and irrationalism. In: *Consequences of pragmatism*. Minneapolis (MN): University of Minnesota Press. 160-175. First printed in: 1980. *Proceedings of the American Philosophical Association*. 53(6):719-738.
- Rorty R. 1982c. The world well lost. In: *Consequences of pragmatism*. Minneapolis (MN): University of Minnesota Press. 3-18. First printed in: 1972. *Journal of philosophy*. 69(19):649-665.
- Rorty R. 1989. *Contingency, irony, and solidarity*. Cambridge (UK): Cambridge University Press.
- Rorty R. 1991a. Science as solidarity. In: *Objectivity, relativism, and truth: philosophical papers, vol. 1*. Cambridge (UK): Cambridge University Press. 35-46. First printed in: 1987. Nelson JS, Megill A, McCloskey DN, editors. *The rhetoric of the human sciences*. Madison (WI): University of Wisconsin Press. 38-52.
- Rorty R. 1991b. Texts and lumps. In: *Objectivity, relativism, and truth: philosophical papers, vol. 1*. Cambridge (UK): Cambridge University Press. 78-92. First printed in: 1985. *New literary history*. 17(1):1-16.
- Rorty R. 1991c. Pragmatism, Davidson, and truth. In: *Objectivity, relativism, and truth: philosophical papers, vol. 1*. Cambridge (UK): Cambridge University Press. 126-150. First printed in: 1986. LePore E, editor. *Truth and interpretation: perspectives on the philosophy of Donald Davidson*. Oxford (UK): Blackwell. 333-355.

- Rorty R. 1991d. The priority of democracy to philosophy. In: Objectivity, relativism, and truth: philosophical papers, vol. 1. Cambridge (UK): Cambridge University Press. 175-196. First printed in: 1988. Peterson MD and Vaughn RC, editors. The Virginia statute for religious freedom. Cambridge (UK): Cambridge University Press. 257-282.
- Rorty R. 1998a. Human rights, rationality, and sentimentality. In: Truth and progress: philosophical papers, vol. 3. Cambridge (UK): Cambridge University Press. 167-185. First printed in: 1993. Shute S, Hurley S, editors. On human rights: the 1993 Oxford Amnesty Lectures. New York (NY): Basic Books. 111-134.
- Rorty R. 1998b. Feminism and pragmatism. In: Truth and progress: philosophical papers, vol. 3. Cambridge (UK): Cambridge University Press. 202-227. First printed in: 1990. Tanner Lectures on Human Values. 1-35.
- Scanlon TM. 1998. What we owe to each other. Cambridge (MA): Harvard University Press.
- Schwartz R. 2016. Pragmatic constructivism: values, norms, and obligations. In: Marchetti G, Marchetti S, editors. Facts and values: the ethics and metaphysics of normativity. New York (NY): Routledge. 42-59.
- Shafer-Landau R. 2003. Moral realism: a defence. Oxford (UK): Oxford University Press.
- Shklar J. 1984. Putting cruelty first. In: Ordinary vices. Cambridge (MA): Harvard University Press. 7-44.
- Singer P. 1971. Famine, affluence, and morality. *Philosophy & public affairs*. 1(3):227-243.
- Sinnott-Armstrong W. 2008. Framing moral intuitions. In: Sinnott-Armstrong W, editor. *Moral psychology, volume 2: the cognitive science of morality: intuition and diversity*. Cambridge (MA): Cambridge University Press. 47-76.
- Sinnott-Armstrong W. 2019. Consequentialism. In: Zalta EN, editor. *Stanford encyclopedia of philosophy*. Summer 2019 edition. <<https://plato.stanford.edu/entries/consequentialism/>>.
- Sinnott-Armstrong W, Young L, Cushman F. 2010. Moral intuitions. In: Doris J, Moral Psychology Research Group, editors. *The moral psychology handbook*. Oxford (UK): Oxford University Press.
- Smilansky S. 2003. Can deontologists be moderate? *Utilitas*. 15(1):71-75.
- Sosa E. 2007. *A virtue epistemology: apt belief and reflective knowledge, volume i*. Oxford (UK): Oxford University Press.
- Street S. 2008. Constructivism about reasons. In: Shafer-Landau R, editor. *Oxford studies in metaethics, volume 3*. Oxford (UK): Oxford University Press. 207-246.

- Thomson JJ. 1985. The trolley problem. *Yale law journal*. 94(6):1395-1415.
- Väyrynen P. 2021. Thick ethical concepts. In: Zalta EN, editor. *Stanford encyclopedia of philosophy*. Spring 2021 edition. <<https://plato.stanford.edu/entries/thick-ethical-concepts/>>.
- West C. 1989. *The American evasion of philosophy: a genealogy of pragmatism*. Madison (WI): University of Wisconsin Press.
- Walker MU. 1996. Feminist skepticism, authority, and transparency. In: Sinnott-Armstrong W and Timmons M, editors. *Moral knowledge? New readings in moral epistemology*. Oxford(UK): Oxford University Press. 267-293.
- Williams B. 1987. *Ethics and the limits of philosophy*. Cambridge (MA): Harvard University Press.
- Williams B. 2002. *Truth and truthfulness: an essay in genealogy*. Princeton (NJ): Princeton University Press.