

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Rethinking the objectivity of ethics

Permalink

<https://escholarship.org/uc/item/98w94220>

Author

Kurth, Charles Burke

Publication Date

2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Rethinking the Objectivity of Ethics

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Philosophy

by

Charles Burke Kurth, Jr.

Committee in charge:

Professor David Brink, Chair
Professor Richard Arneson
Professor Jonathan Cohen
Professor Alan Houston
Professor Andrew Kehler
Professor Gila Sher

2011

Copyright

Charles Burke Kurth, Jr., 2011

All rights reserved.

The Dissertation of Charles Burke Kurth, Jr. is approved, and it is acceptable
in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2011

For my father.

TABLE OF CONTENTS

Signature Page	iii
Dedication.....	iv
Table of Contents	v
List of Tables and Diagrams.....	vii
Acknowledgments	ix
Curriculum Vita.....	xi
Abstract.....	xii
Chapter 1. A Puzzle, a Project.....	1
1. Dimensions of Objectivity: Some Distinctions	3
1.1. The Semantic Dimension	3
1.2. The Ontological Dimension	5
1.3. The Metaphysical Dimension.....	6
1.4. An Initial Characterization	7
2. Can Our Objectivity Talk be Vindicated?	9
2.1. The Case for Factualism.....	9
2.2. The Case Against Moral Realism.....	10
2.2.1. The Argument from Prescriptivity	11
2.2.2. The Argument from Inquiry and Disagreement	12
2.3. Does Skepticism about Objectivity Loom?	14
2.4. The Factualists' Challenge	16
3. The Expressivists' Insight	17
3.1. The Expressivists' Strategy	18
3.2. An Enriched Array of Options	21
4. What Success Would Look Like	24
4.1. The Marks.....	24
4.2. The Status of the Marks.....	26
4.3. From Marks to Success Conditions	28
5. The Plan of Attack	29
Chapter 2. Logic From Morals, Morals from Logic.....	31
1. Background: Logical Consequence and Logical From	35
1.1. The Importance of Logical Form	35
1.2. The Modal Test.....	37
2. Meta-Ethics and the Modal Test.....	38
3. The Limits of Expressivism.....	41

3.1. Appeals to the Nature of Thought and Language: Version 1	42
3.2. Appeals to the Nature of Thought and Language: Version 2	46
3.3. Appeals to an Extra-Mental Account of Formality	51
3.4. The Verdict	58
4. Might Expressivists Gain from a Skeptical Turn?	59
5. Conclusion	62
Chapter 3. Realism and Moral Error	64
1. Explaining Error: A Second Test	65
2. The Independence of Moral Properties and Facts	68
3. Identity Strategies	70
3.1. Explicit A Priori Definitions	71
3.2. Implicit A Priori Definitions	75
3.3. Unanalyzable, Sui Generis Primitives	79
4. Functionalist Strategies	82
4.1. What Moral Functionalism Is	82
4.2. The Functionalists' Critical Argument	85
4.3. The Functionalists' Positive Case: The Analogy Argument	88
4.4. The Functionalists' Dilemma	92
4.5. The Challenge from Fashion	93
4.6. The Search for a More Substantive Standard	96
4.6.1. Belief Independent Explanation	97
4.6.2. Deep Error	100
4.7. The Verdict	103
5. Conclusion and Implications	104
Chapter 4. Constructivism and Moral Error	106
1. Sentimentalist Views	107
2. Idealization Views	115
3. Rationalist Views	120
3.1. Agency Accounts	120
3.2. Hypothetical Agreement Accounts	129
3.2.1. Scanlon and Moralized Agreement	130
3.2.2. Gauthier and Non-Moralized Agreement	133
4. The Verdict on Constructivism	137
Chapter 5. Transforming Moral Conventionalism	140
1. Moral Conventionalism—First Steps	142
1.1. Conventions and Conventionalism	142
1.2. Moral Conventionalism	143
1.3. Moral Conventionalism, Subjectivism, & Relativism	150
2. The Allure of Moral Conventionalism	151
3. Conventionalism and Moral Error	153
3.1. The Need to Focus on “Corrected” Conventions	153
3.2. The Problem of Specifying “Corrected” Conventions	156

3.2.1. Non-Hypothetical Strategies	157
3.2.2. Hypothetical Strategies.....	158
3.3. Why this is a Problem Conventionalists Cannot Accept.....	160
4. The Subjectivist Turn	163
Chapter 6. Sophisticated Subjectivism	165
1. Orientations	166
2. The Place of Moral Conventions	167
3. The Nature and Purpose of Moral Judgment.....	169
4. Validation: Its Nature and Function	179
4.1. Dimensions of Validation: Truth and Empathy.....	176
4.2. Structural Challenges.....	177
4.3. Subjective Validation	180
5. A Procedure for Fixing the Moral Facts.....	182
5.1. The Proposal	182
5.1.1. The Truth Requirement	183
5.1.2. The Empathy Requirement.....	185
5.2. Objections and Replies	188
5.2.1. Metaphysical Austerity Lost.....	188
5.2.2. Circularity.....	189
5.2.3. Genuine Disagreement	190
5.3. The Verdict.....	191
6. Fallibility	192
7. Normativity.....	197
7.1. Universality	198
7.2. Categorical.....	199
7.3. Authority & Stability	199
8. A Better Account of Moral Facts	201
References	205

LIST OF TABLES AND DIAGRAMS

Table 1: An Initial Characterization	7
Table 2: A Revised Characterization.....	23
Diagram 1: A Map of Logical Space.....	24

ACKNOWLEDGMENTS

This has been a quite a journey. When I first considered pursuing a doctorate in philosophy while I was working in finance in New York City, I had no idea what I'd be getting myself into—or how enjoyable it would be.

Many thanks are due to those who helped along the way. I owe much to Brad Armendt, Stewart Cohen, John Devlin, and Bernie Kobes at Arizona State—I learned much from them while pursuing my M.A. degree there. A larger debt is owed to those at UCSD whose insights and encouragement have helped me write this dissertation: To Dick Arneson, Jonathan Cohen, Gila Sher, and—especially—David Brink. Thanks as well to Alan Houston and Andy Kehler for their questions and their time.

Thanks are also owed to many others who, through comments and discussion, have helped me develop various parts of this project: Paul Bloomfield, Justin D'Arms, Dale Dorsey, Jamie Dreier, Jyl Gentzler, Shaun Nichols, Howard Nye, Mark Schroeder, and David Wong. I am especially grateful to Nina Brewer-Davis and Mike Tiboris—our weekly meetings to talk philosophy were invaluable.

Finally, I thank my family for their support. And to Pippin Schupbach, thanks for your love and encouragement—and for making me laugh.

Chapter 2 is a reprint, with slight modifications, of “Logic from Morals, Morals from Logic,” *Philosophical Studies*, forthcoming. Permission to reproduce this material has been granted by the copyright owner. The dissertation author was the primary investigator and author of this paper.

Chapter 4 contains material that has been submitted for publication and may appear in *Philosophical Studies*. The dissertation author was the primary investigator and author of this paper.

Chapters 4, 5, and 6 contain material that is being prepared for submission for publication. The dissertation author was the primary investigator and author of this work.

CURRICULUM VITA

Education

- 2011 University of California, San Diego
 Ph.D. in Philosophy
- 2003 Arizona State University
 M.A. in Philosophy
- 1997 University of Virginia
 B.A. in Economics, B.A. in U.S. Government

Publications

- Kurth, Charlie. (forthcoming). "Logic for Morals, Morals from Logic."
Philosophical Studies.

ABSTRACT OF THE DISSERTATION

Rethinking the Objectivity of Ethics

by

Charles Burke Kurth, Jr.

Doctor of Philosophy in Philosophy

University of California, San Diego, 2011

Professor David Brink, Chair

Commonsense says that morality is objective. But skeptics deny that the normativity and fallibility that characterize this objectivity really exist. I disagree. I develop an account of moral facts that vindicates the objectivity of ethics. Moreover, I do this without relying on either the strong foundational claims of realism and constructivism, or the revisionary logic of expressivism. Instead, I build from a moral psychology that is modest and familiar—one that focuses on the distinctive, but well-

known, anxiety that is provoked when we make a moral judgment.

Drawing on empirical work, I argue that this anxiety is a universal feature of our social-psychological lives, one we've developed to promote social cohesion. Moral anxiety does this by altering our orientation to others: we become more sensitive to the consequences that our actions can have, and so tend to avoid them if we find that they are not justified. I argue further that this picture of the role that moral anxiety serves points to a broadly teleological account of moral facts: moral facts are the upshot of a validation procedure that alleviates moral anxiety in a manner consistent with the function that it serves.

I then develop an account of the form of validity that is constitutive of moral facts. I argue that because moral anxiety is provoked by unease about the legitimacy of one's moral judgment, it will be properly alleviated only if that judgment (i) is not based on false non-moral beliefs, and (ii) uses principles that correctly assess the situations of those affected by the conflict in question. I argue that what counts as 'properly assessing the situation of the affected parties' is subjectively fixed in the sense that it is the upshot of what one would defend, on reflection and when pressed, given a motivation to resolve moral conflict in a manner that preserves social cohesion. Despite its subjective nature, my proposal captures the objectivity of ethics: fallibility is possible since the judgments we actually make can come apart from those that we would defend; normativity is explained by the sensitizing effects of moral anxiety.

Chapter 1: A Puzzle, a Project

Philosophers and non-philosophers alike speak about the objectivity of ethics. This talk seems plausible in light of what we think and say about moral matters. For instance, the moral judgment ‘torture is wrong’ is like non-moral judgments like ‘stop signs are red’ and ‘electrons are negatively charged’ in that we take both to be making true assertions about the way the world is. Furthermore, when faced with a difficult moral question, we engage in deliberation and inquiry in the hope of coming to a principled answer to the issue at hand. Moreover, we also presume that the moral judgments we make are susceptible to correction and improvement. In fact, the mere presence of such critical practices suggests morality admits of fallibility; it suggests that there is something we are aiming to get right. These and other features appear to confirm that the common thought that moral discourse functions to make claims about an objective realm of properties and facts.

However, this is only part of the picture. Other features of our ethical practice call into question the plausibility of our “objectivity” talk. For instance, further reflection reveals reasons to doubt the observations of the previous paragraph. For starters, the claim that moral judgments are fact stating is hard to square with the observation that moral judgment—unlike, say, judgments about electrons—seems essentially prescriptive. Consider: only in the case of a moral judgment like ‘torture is morally reprehensible’ does the *mere* act of making the judgment seem to motivate one to act as in certain ways. This suggests that moral judgments do not function to describe. Rather, it’s more plausible to maintain that they function to express one’s

subjective attitudes—one's strong distaste for torture, perhaps. We get comparable reasons for doubt regarding the above claim that moral inquiry is a process of discovery. After all, moral inquiry is typically an introspective process, not one involving observation or experiment. This suggests that it is better understood as a refining of (say) the coherence of our conventions or attitudes than a search for some realm of objective moral facts. For reasons like these, many are led to skepticism about the objectivity of ethics.

This conflicting data leaves us with a puzzle: Can we substantiate the objectivity of ethics? More specifically, suppose—as most do—that the features noted in the first paragraph provide a rough picture of the objectivity of ethics. Is there a plausible rendering of the details of these features, one that would explain the way in which ethics is objective? Furthermore, could we provide a philosophical foundation for such an objectivist picture, one capable of turning back skeptical doubts about the legitimacy of our objectivity talk?

In this thesis, I argue that ethics is objective. But I also maintain that existing accounts cannot adequately capture this—they rely on assumptions that are either implausibly strong or excessively revisionary. The alternative that I develop grounds the objectivity of ethics in familiar features of our psychology. The result is a more sophisticated picture of the distinctive dimensions of moral objectivity. I begin, in this chapter, by providing a sharper picture of the project at hand. To accomplish this I set three goals: (1) Explain why providing an account of the objectivity of ethics is a formidable, but important challenge. (2) Identify both the features of our moral

discourse that make it seem objective, and situate existing views with regard to their understanding of those features. (3) Propose a strategy for investigating the objectivity of ethics, and articulate a set of success criteria that any viable account must meet.

1. Dimensions of Objectivity: Some Distinctions

Providing a definition or even proposing a framework for investigating questions of objectivity is messy and controversial work.¹ But having a framework is necessary if inquiry is to proceed. In what follows, I identify a set of semantic, ontological, and metaphysical claims that provide one plausible way of making sense of what the objectivity of ethics might consist in.² With these distinctions in place, we will be able to both formulate an initial characterization of the ways in which one might make sense of the objectivity of ethics, and begin to understand why vindicating our objectivity talk is a significant challenge.

(1.1) The Semantic Dimension. The features of our moral discourse and practice noted above suggest at least a couple ways in which ethics might be objective. With regard to its semantic or linguistic function, ethics might be “objective” in the sense that moral judgments aim to state facts about some aspect of the world (whether

¹ See Brink (1989: 14-22), Shafer-Landau (2003: 13-8), Sober (1982), Wong (1984: chap. 1), and Wright (1988).

² Though I will not argue for it here, I believe the framework that follows to be superior to the prominent alternatives. First, the “Moral Problem” framework advocated by Smith (1995) and others is problematic because of its heavy reliance on a confused conception of moral motivation (see Kurth ms. 1 for details). Second, the difficulty with views that understand questions of objectivity in terms of questions about the nature of the truth predicate (e.g., Wright 1992) is that they depend on an implausible framework for understanding questions about the nature of truth (this issue will be briefly discussed in chap. 2).

there are such facts is a further *metaphysical* question—more on this below).³ Call this conception of the objectivity of ethics Factualism. More specifically, the claim is this:

Factualism: Moral judgments are objective in that they express content (e.g., propositions, states of mind) that seeks—in some way—to make claims about a distinctive realm of moral facts.⁴

There are two ways to reject Factualism. First, one might deny that moral judgments express any sort of content. But this is clearly absurd; it entails that moral judgments are nothing more than gibberish. The alternative that many have found compelling maintains that, though moral judgments express content, that content no more makes a claim about some aspect of the world than does a groan of pain. Rather, moral judgments like ‘torture is wrong’ are akin to an exclamation of ‘that’s hysterical!’ on hearing a witty joke, or an imperative like ‘don’t walk on the grass!’ Those who reject Factualism maintain that the linguistic function of moral judgments is not to make claims about facts, but rather to vent things like feelings, tastes, preferences, or commands.

³ Unless otherwise noted, I use ‘judgment’ (of whatever type—moral, scientific, prudential...) to pick-out a particular type of act unique to the discourse in question. I do not attach any metaphysical (or other) substance to it. So, for instance, I intend ‘moral judgment’ to be neutral between different meta-ethical accounts.

⁴ Factualism is often combined with, or seen as being the same as, cognitivism. Cognitivism is the view that moral judgments express a belief-like mental state. It contrasts with non-cognitivism—the position that takes moral judgments to express conative mental states like desires. Moreover, on the traditional picture the cognitive and non-cognitive options are seen as mutually exclusive and exhaustive. But meta-ethicists have increasingly moved to either deflate the notion of a belief so that it can have conative force (e.g., Blackburn 1984, 1998; Gibbard 2003; Horgan and Timmons 2006; Horwich 1998; Wright 1992), or claim that moral judgments can express a complex mental state that is both belief-like and conative (e.g., McDowell 1985, Wiggins 1987). The result is that the cognitivist/non-cognitivist distinction is increasingly unhelpful.

(1.2) The Ontological Dimension. But even were we to grant that ethics is objective in the Factualist sense, we might still wonder whether it is “objective” at a deeper level. More specifically, if moral judgments seek to represent moral facts, then they function to attribute moral properties (e.g., wrongness, goodness) or relations (e.g., morally better than) to acts and individuals. Noticing this points to a set of questions about the ontological nature of these moral properties and relations. For example, we might wonder what sort of connection exists between these moral properties, relations, and facts on the one hand, and our thoughts (cognitive and conative), institutions (social, cultural, religious, etc.), or evidence (*a priori* or *a posteriori*). Is the moral constructed from these things, or is it independent of them? From this question we can specify a second dimension of objectivity, one concerning the ontological nature of moral properties and relations:⁵

Independence: Moral judgments are objective in the sense that if the moral properties and relations that they seek to pick out (and so the moral facts that they purport to represent) exist, they do so independently of our thoughts, institutions, or evidence.⁶

To accept or reject Independence is to take a stand on what it would take for moral properties or facts to exist.⁷ Consider some analogies. If moral judgments like ‘murder is wrong’ are Independent, then their ontological nature is like that of our

⁵ Ontological questions of this sort are often combined with questions of naturalism—e.g., are moral properties naturalistic in the sense of falling, in some way, within the scope of physical and social sciences? I think putting much initial weight on considerations of naturalism is a mistake. For one, there is very little consensus about what it means for something to be “naturalistic” and so there is no way to frame the issue of naturalism that is not highly controversial. But, as we will see, this does not mean that questions of naturalism are irrelevant to ontological issues about the objectivity of ethics.

⁶ I offer a more refined account of Independence in Kurth ms. 2. C.f., Shafer-Landau 2003: 13-8.

⁷ Importantly, we do not (yet) have a metaphysical commitment to there *actually* being such properties or facts. That further dimension of objectivity will be discussed below.

commonsense conception of scientific judgments like ‘electrons have negative charge’ in that the nature of the properties that these judgments pick-out (wrongness, negative charge) is not a function of our thoughts, institutions, or practices. By contrast, if moral judgments are not Independent, then their ontological nature is more like our common understandings of other types of judgments. For instance, they might be like (i) judgments about preferences (e.g., ‘chocolate is better than vanilla’) in being based on our desires, or (ii) judgments about etiquette (e.g., ‘napkins are placed on the left’) in being dependent on our (social) institutions.

(1.3) The Metaphysical Dimension. But once we have a conception of the ontological nature of the moral properties and relations that advocates of Factualism take moral judgments to involve, we can ask a further metaphysical question—Are there such things? Those who answer negatively endorse an error theory about our moral discourse. They maintain that moral judgments are pervasively and systematically misleading in that they purport to be about a realm of properties of facts—be they Independent or not—that does not actually exist. Thus talk of morals is like talk of witches or phlogiston. By contrast, those who maintain that there are moral properties and facts (again, Independent or not) endorse some version of Realism. Thus we have a third dimension of objectivity:

Realism: Moral judgments are objective in the sense that the moral properties, relations, and facts that they seek to pick out (be they Independent or not) actually exist.⁸

⁸ This notion of ‘Realism’ is broader than some uses of the term in that it encompasses both versions of Factualism that accept Independence *and* those that do not. In what follows, I will use ‘Realism’ to refer

(1.4) An Initial Characterization. Given these three dimensions of objectivity we can both partition logical space into three broad groups, and locate some standard meta-ethical positions therein (this initial characterization will be refined in what follows):

Table 1: An Initial Characterization⁹

Semantically Factual?	Ontologically Independent?	Metaphysically Real?	
		Yes	No
Yes	Yes	Analytic naturalism, non-reductive naturalism, non-naturalism, divine command theories...	Error theories
Yes	No	Practical reason theories, conventionalism, response-dependence theories...	Error theories
No	No	N/a	Emotivism, prescriptivism

Let me say a few further words about Table 1. The non-error theoretic views of the first row—the moral realist accounts—see ethics as both Factualist and Independent. Views of this sort differ over *inter alia* how we are to understand the nature of the independence of moral properties (e.g., is the independence of moral

to the broader notion, and ‘realism’ to refer to the narrower use that only refers to Factualist accounts that endorse Independence.

⁹ Non-Factual Independence is not a coherent option.

properties a result of their being natural properties? Supernatural ones? *Sui generis* ones?...). As we will see (in chap. 3), these differences often result from different conceptions of what it means for moral facts to be ontologically independent. The non-error theoretic views of the second row—the constructivist accounts—share the realists' commitment to Factualism, but deny the commitment to Independence. Again, a wide range of views occupies this space. These views differ about both why ethics fails to be Independent, and on what sort of things moral judgments depend. The final row—non-cognitivism—denies that ethics is either Factualist or Independent.¹⁰ A final point: Given this mapping of logical space, we can distinguish accounts that are in some way skeptical about the prospects of vindicating our objectivity talk (i.e., error theories and all varieties of non-cognitivism), and those that are not (i.e., all non-error theoretic versions of Factualism).

Which of these models provides the best fit? I will argue that moral judgments are fact stating. But these facts are about a rich and complicated aspect of our lives—an aspect that involves both our thoughts and attitudes about how to adjudicate moral conflict. Furthermore, I maintain that these facts correspond to properties that actually exist. Thus, I will defend a non-error theoretic conception of moral constructivism. But, as will become apparent in the chapters to follow, I find none of the existing views satisfactory. I maintain that seeing why points to a new and better alternative.

¹⁰ As will become more apparent below, I use 'non-cognitivism' to refer to emotivism (e.g., Ayer 1936/1946, Stevenson 1937) and prescriptivism (e.g., Hare 1952). It excludes the meta-ethical expressivism found in, for instance, Blackburn 1984, 1993, 1998; Gibbard 1990, 2003; and Horgan & Timmons 2006. Expressivism will be discussed in §3 below and at length in chapter 2.

2. Can Our Objectivity Talk be Vindicated?

If the goal is to defend a version of constructivism, the above discussion suggests an obvious strategy: Start by demonstrating that our moral discourse is Factualist, and then argue that it is not Independent. The result would vindicate moral constructivism over moral realism. But this is easier said than done. As will become apparent, it is not easy to deny Independence without also having to either (i) deny Factualism (and so become an advocate of non-cognitivism), or (ii) deny that there are any moral properties and facts (thereby falling into an error theory). Having to concede one of these would seem to entail conceding that we cannot vindicate our objectivity talk. The rest of this section attempts to make this challenge explicit.¹¹

(2.1) The Case for Factualism. In order to begin to understand why Factualism seems true of our moral discourse, consider some typical moral judgments:

- (1) Philanthropy is good.
- (2) It's false that the invasion of Iraq was just, but it is true that Saddam was morally depraved.
- (3) If stem-cell research is wrong, abortion is too.
- (4) George believes that stem-cell research is wrong.
- (5) Nancy acted rightly; therefore, someone acted rightly.
- (6) Murder is wrong; therefore, hiring a hitman is wrong.

Simple moral judgments like (1) have declarative surface syntax and are expressed in the indicative mood. (2) demonstrates that we use moral judgments to make assertions that we view as being truth-apt. In fact, it seems that we take many of our moral

¹¹ Note, I do not take the arguments that follow to be conclusive. Rather, I draw on a set of considerations that have historically played an important role in framing debates about the objectivity of ethics.

judgments to be true. Furthermore, the syntactic form of moral judgments allows them to be embedded in truth-functional constructions (e.g., the conditionals and conjunctions of (2) and (3)), propositional attitude ascriptions (the belief ascription of (4) and any other attitude that could take the place of ‘believes’), and ‘that’ clauses (e.g., (2)). Moral judgments also seem to figure in legitimate and illegitimate inferences and arguments—be these “deductive” like in (5), or “inductive” like (6). This data strongly suggests that moral judgments are in the business of stating facts. The case for Factualism gains further plausibility from the observation that there are few areas of discourse that display similar logical and syntactic features, but where we nonetheless take these features to be misleading.

Historically, considerations like these have been taken to warrant the conclusion that the function of moral judgments is to state facts. David Copp expresses this point nicely:

the form and content of our moral [judgments] and the nature of moral deliberation and argument can most naturally be explained on the assumption that these [judgments] are beliefs in the ordinary sense of the term and that moral judgments are, or express, propositions. (1991: 613; c.f., Brink 1989: 23-31; Mackie 1977: 30-5; Railton 1989)

While some like Copp see this evidence as decisive, others take them as effectively shifting the burden of proof to the Non-Factualists. Either way, we have good reasons to endorse Factualism.

(2.2) The Case Against Moral Realism. Arguments against moral realism often begin with the plausible claim that if our moral discourse is fact stating, then it must be *about* the moral features of the world. That is, moral discourse serves to *describe* a

realm of moral properties and facts. These arguments then note significant differences between moral properties and facts, and the properties and facts of paradigmatically descriptive discourses (e.g., middle-sized dry goods, science).¹² Typically, moral discourse is thought different for one of two reasons: (i) because moral judgments are essentially prescriptive, and (ii) because of the distinctive nature of moral inquiry and disagreement. Regardless of which difference is appealed to, the conclusion that results is the same: Moral properties and facts are best understood as human constructions rather than independent entities. Let's look at each of these arguments more closely.

(2.2.1) The Argument from Prescriptivity. Almost all meta-ethicists agree that a distinctive, seemingly fundamental feature of our moral discourse is its prescriptive force.¹³ More specifically, moral judgments are seen as providing us with a reason or a motivation to act in particular way: My judging that I ought to give to charity seems to give me at least some reason or motivation for making philanthropic contributions. In this way, moral judgments appear to be quite different from judgments about paradigmatically mind-independent areas of discourse like science. For notice, judging that electrons have charge does not—at least on its own—incline one toward any specific sort of behavior. Thus, we should reject moral realism in favor of some form of constructivism. Because we have already made a presumptive case for Factualism,

¹² Reasoning of this sort is offered by Blackburn (1998: 86ff), Darwall, Gibbard, & Railton (1992: 130), Mackie (1977: 38-42), Lillehammer (2007), and others.

¹³ A representative sample: Blackburn 1998, Brink 1989, Mackie 1977, McDowell 1985, Scalton 1998, and Smith 1995.

this asymmetry seems to warrant the conclusion that the properties picked out by our moral judgments must be tied to our beliefs, desires, or institutions.¹⁴

(2.2.2) The Argument from Inquiry and Disagreement. But even if we set the prescriptive or action-guiding features of moral judgment aside, there are other considerations that make it difficult to understand how moral properties and facts could be Independent. More specifically, looking to certain features of the critical practices we employ when we think and talk about moral matters makes it hard to see how an advocate of moral realism can provide a plausible account of either what these properties and facts are, or how we might come to know anything about them.

Consider, for instance, one dimension of moral disagreement.¹⁵ Almost all meta-ethicists acknowledge that our endorsement of a particular moral judgment is a function of the justification we take there to be for that judgment—that is, moral judgments purport to be responsive to the reasons we take to support them. In light of this reasons responsiveness, if we are to make sense of moral disagreement as *reasoned* disagreement in this sense, we must be presupposing that there is an intimate connection between the moral facts (e.g., that torture is wrong) and the non-moral ones (e.g., that torture causes pain).¹⁶ To see why, notice that when we're engaged in

¹⁴ See, for instance, Korsgaard 2009 and Smith 1995. Railton 1986a, 1986b, 1989 also seems to endorse an argument of roughly this form—at least for our conception of what is intrinsically valuable for a person.

¹⁵ There is an extensive literature on what moral disagreement is and what it presupposes. The discussion that follows touches on just one small dimension of this debate; further aspects will be discussed in chap. 3.

¹⁶ This point is actually more general—the intimate relation seems to hold not just for the moral, but for the normative/evaluative more broadly. See Hare 1952: chap. 8.

moral disagreement, we appeal to reasons for our conclusions and against the conclusion of our opponent. But—importantly—these reasons often involve *non-moral* considerations (e.g., that torture causes pain). Since we take these non-moral considerations *to be reasons* for moral judgments, we must think that there is some intimate relation between the moral question at issue, and the non-moral considerations we appeal to. The challenge then is to provide an account of this intimate connection.

One way to do this would be to hold that moral facts just are subjective responses to non-moral facts. While this line is compatible with constructivism, a realist must, for obvious reasons, reject it. In order to eliminate an appeal to our subjective responses, realists have several options—but none seem satisfactory. First, they can attempt to analyze the relation between the moral and the non-moral. Is this an identity or analytic relation? If so, then isn't the assumption of *moral* facts otiose? If not, what exactly is the relation? Plausible answers are notoriously hard to come by.¹⁷ Second, they might take the relation as primitive—they could take these moral/non-moral relations to be brute, unexplainable features of the world. But such a move makes it very difficult to give a plausible account of our epistemological and critical practices. For instance, it makes it very difficult to understand what justifies our claims about the relations between the moral and the non-moral—say, that it's the

¹⁷ A sampling of arguments against the possibility of providing an account of this intimate connection can be found in Blackburn 1984: 187-9, Darwall, Gibbard, & Railton 1992: 172-4, 177-8, Johnston 1989, Mackie 1977: chap. 1, Moore 1903, and Timmons 1999: chaps. 1-2. I will discuss aspects of this debate further in chap. 3.

painfulness of torture that makes it wrong.¹⁸ Finally, the realist might maintain that while reductive analyses of the relation aren't possible, our claims about these relations can nonetheless be justified. Not only are such claims intuitively plausible, but they are supported by our reflections and observations. But many doubt the sufficiency of this type of justification—either because it's too vague, or because it's too difficult to reconcile with other features of our moral discourse (e.g., the pervasiveness of moral disagreement).¹⁹ So we seem to have another reason to reject Independence and advocate some form of constructivism.

(2.3) Does Skepticism about Objectivity Loom? Unfortunately, these two arguments against moral realism afford constructivism more plausibility than it seems to deserve; buying into either of them seems to require rejecting not just realism, but Factualism as well. The reason is that these arguments readily generalize to show that constructivism is in no better a position to explain the prescriptive, epistemological, and critical aspects of our moral discourse and practice than is realism.

Recall that the argument from prescriptivity concluded that realism appears unable to make sense of the fact that moral judgments seem to be essentially prescriptive. But as Simon Blackburn notes, this is a difficulty that also afflicts constructivism:

[T]he fact [is] that valuing and other ethical activities are different from

¹⁸ See Darwall, Gibbard, & Railton 1992: 162-4, Hare 1981, Harman 1977: chap. 1, and Mackie 1977: chap. 1 for representative arguments. This problem will also be pursued further in chap. 3.

¹⁹ For arguments that the assumption that a non-reductive, non-primitive connection is not sufficient see, for instance, Leiter 2001, Lillehammer 2007, Nichols 2004: 149-64, and Timmons 1990. Again, this matter will receive further attention in chap. 3.

describing and explaining, or from purely scientific representations. ... [I]t should have been obvious from the beginning that since moralizing and valuing are distinct activities, the words we use to communicate our morals and our values will have their distinctive meanings. Why try to square the circle by going bull-headed for an illuminating ‘account’ of their ‘truth conditions’? (1998: 86-7)²⁰

The constructivist tries to accommodate the prescriptive aspects of moral judgments by maintaining that moral properties and facts are a function of things like our beliefs, desires, and institutions. What Blackburn notes though, is that the constructivists’ move only seems to allow them to provide an account of the content of a moral judgment is. Yet securing such an account only provides something *descriptive*. So constructivists have not managed to capture the *prescriptive* role that is distinctive of moral judgments. The upshot is that to think that *any* form of Factualism can be retained once one acknowledges the distinctive prescriptive role of moral judgments is just plain bull-headed—for there’s nothing but dogmatism left to support it.

Turning to the second argument against realism, it seems that the constructivist is in no better a position to explain the reasons responsiveness of moral judgments than was the realist. More specifically, the constructivists’ move to make sense of reasons responsiveness in terms of something like our subjective responses seems merely to trade one puzzle for another. The move to explain the intimate connection between the moral and the non-moral (and thereby explain the reasons responsiveness of moral judgments) is only helpful to the extent that we have reason to think that our

²⁰ Similar lines of thought are expressed by Blackburn (1984: 188); Darwall, Gibbard, Railton (1992: 137, 143); Hare (1952: chap. 7); Gibbard (1990: chap. 1); Timmons (1999: 21, 39-40).

subjective responses can justify moral judgments. But given the commonsense thought that the legitimacy of a moral judgment is a function of (among other things) its *independence* from our subjective responses, few find this argument compelling.²¹ Thus endorsing the above arguments against realism seems to push one toward skepticism about the objectivity of ethics. The form that one's skepticism takes depends on how one understands the force of these arguments—Do they undermine the existence of moral properties and facts (error theory),²² or do they undermine Factualism directly (non-cognitivism)?²³

(2.4) The Factualists' Challenge. So we arrive at the challenge: Our moral talk displays a broad set of logical and syntactic features, features that strongly suggest that ethics is objective in at least the sense that some non-error theoretic version of Factualism seems true. But our moral discourse and practice displays other features—epistemological, critical, and prescriptive—that raise doubts about whether ethics exhibits the objectivity of realism. But these features also call into question whether ethics could even be Factualist and so calls into question whether it is possible to vindicate the objectivity that we intuitively take our moral discourse and practice to possess.

²¹ Arguments of roughly this sort can be found in Darwall, Gibbard, & Railton 1992: 143-4. See Korsgaard (1996: 45-6) also provides an argument that one can't reject Independence without also rejecting Factualism (note that Korsgaard does not endorse this line).

²² On at least one reading, this is strategy employed in Mackie (1977: chap. 1).

²³ This strategy is most clearly advocated by Ayer (1936/1946: chap. 6). One might also read Stevenson (1937) and Hare (1952) as proffering similar lines.

However, endorsing skepticism of an error theoretic or non-cognitivist variety requires one to assume a serious cost in terms of plausibility. One has to admit either that *all* of our moral judgments are false, or that the surface syntax and logical form of our moral language is *grossly* misleading—initial appearances to the contrary, moral judgments just aren't in the business of stating facts. So there are good general methodological reasons—reasons independent of a commitment to constructivism—to prefer a non-skeptical solution. Because of this, we should try to avoid skepticism. But how might one substantiate the objectivity of ethics?

3. The Expressivists' Insight

The non-error theoretic varieties of Factualism noted in Table 1 are not the only meta-ethical views that are concerned with the skeptical threat. A fairly new meta-ethical view—expressivism—takes seriously the metaphysical concerns of the error theorists and the non-cognitivists, but does so in a way that does not entail the high cost of either rejecting Factualism, or insisting that our moral discourse admits of widespread error.²⁴ As we will see, expressivism is significant not just because it reveals a new conception of moral objectivity, but because it also affords alternative accounts with a strategy for avoiding skepticism about the objectivity of ethics. In what follows, I point to a background thesis underlying the skeptics' arguments and show how we can follow expressivists in developing an alternative to it. Doing so will

²⁴ Sophisticated expressivism has its roots in non-cognitivism, but (at least typically) expressivists have been concerned to limit the revisionism that those views required. The affinities between non-cognitivism and expressivism will become more apparent below.

allow us to enrich our conception of the ways in which we might seek to substantiate the objectivity of ethics.

(3.1) The Expressivists' Strategy. Though the expressivist rejects both varieties of skepticism, he sees something right in each view. The expressivist wants to keep the error theorists' commitment to (at least some form of) Factualism, while at the same time retaining the non-cognitivists' claim that the (primary) function of moral judgments is not to make descriptive claims about a realm of moral properties and facts, but rather to express a distinctive set attitudes. More specifically, expressivism seeks to avoid an error theory but nonetheless explain

how we can legitimately say things such as 'It's true that murder is wrong', 'It's false that breaking promises is the right thing to do', 'Jones believes that murder is wrong', and so on, even though we do not *begin* with the assumption that moral predicates refer to properties, that moral judgments express beliefs, or that moral evaluations are truth-apt. It is the project of explaining how we can legitimately talk *as if* we were entitled to the assumption that there is a distinctively moral reality, even though we are not: it is the project of explaining how we can legitimately talk *as if* we are entitled to assume that moral properties express properties, and so on, even though we are not. (Miller 2003: 52, original emphasis)²⁵

But given the framework we have been working with (i.e., that depicted in Table 1), this may not seem a coherent possibility; the expressivists' insight lies in showing us that it is.

Expressivists make a significant contribution to our understanding of the objectivity of ethics by noting that both error theorists and non-cognitivists have

²⁵ C.f., Blackburn 1984: 167-71, 181ff, Gibbard 2003: xii, 62-3, and Horgan & Timmons 2006: 257.

reacted too strongly to a particular thesis about the phenomenology of our moral discourse. Call this thesis Phenomenological Determinism:

Phenomenological Determinism: The phenomenological features of moral judgments are strongly determinate of both (i) the true linguistic function of those judgments, and (ii) the nature/structure of the things those judgments purport to be about.

A few points of elaboration: By “phenomenological features” I’m referring to certain distinctive semantic, critical, and prescriptive marks of our moral discourse and practice—for instance, things like the subject-predicate form of moral judgments, their reasons responsiveness, and the fact that they tend to be action-guiding. What I mean by “strongly determinate...” is that if we endorse Determinism, we are committed the idea that phenomenological features of moral discourse more-or-less literally determine both what the linguistic function of moral judgments is, as well as the underlying nature of the things those judgments are about. Consider an example. Suppose we take the phenomenological features of moral discourse to consist in the features described above. By Determinism, we’re then committed to understanding moral judgments as functioning to predicate properties that are both intrinsically reasons responsive, and action guiding.

In light of Determinism, we can see the expressivists’ as making two related observations. The first notes that both error theorists and non-cognitivists err in reacting too strongly to Determinism. Error theorists endorse it wholeheartedly. For instance, without Determinism, Mackie would not be able to conclude that “in ordinary thought, and even in the meanings of moral terms” our moral judgments function to describe a “peculiar evaluative, prescriptive, intrinsically action-guiding”

realm of properties (1977: 31-2). Non-cognitivists, on the other hand, seem to reject Determinism in full. For instance, they have traditionally claimed that moral judgments behave “logically” in ways that are very different from judgments with seemingly identical syntactic form. But this claim only makes sense on the assumption that they wholly reject Determinism.²⁶

Recognizing that these strong reactions to Determinism were too extreme, we can see the expressivists’ second insight as their staking out of a middle-ground position. Terry Horgan and Mark Timmons, themselves sophisticated expressivists, elaborate this point nicely:

Expressivism thus subsumes old-time versions of noncognitivism. But it leaves open the possibility of a [Factualist] construal of moral thought and discourse. ...[I]t is a metaethical project that embraces an austere irrealist moral metaphysics and yet attempts, in its semantic construal of moral terms and the concepts, to account for the deeply embedded features of moral thought and discourse. (2006: 257)

Expressivists are able to do this not by rejecting Determinism, but rather by softening it—call this moderated position Phenomenological Regulation:

Phenomenological Regulation: The phenomenological features of moral judgments function to constrain—in some way—both (i) the true linguistic function of those judgments, and (ii) the nature/structure of the things those judgments purport to be about.

²⁶ This comes out most clearly in Ayer (1936/46: chap. VI), though the theme is also present to some extent in Hare (1952: chaps. 6-7) and Stevenson (1937: 21-6). The point in the text may be too strong since in various ways non-cognitivists like Hare and Stevenson seem to recognize—at least to some extent—a need to try and accommodate some of the semantic features of moral phenomenology.

This expressivists' insight is significant because opens up a wide range of tools that Factualists can employ in order to resist the skeptics.²⁷ Most significantly, Regulation allows for more flexibility in how a Factualist can understand the interaction between the semantic and ontological dimension of moral judgments.²⁸

(3.2) An Enriched Array of Options. One of the most immediate consequences of trading Determinism for Regulation is that it allows for an expressivist-friendly variety of Factualism—one that can be characterized by its distinctive semantic and ontological claims. Semantically, we have an alternative account of the content expressed by moral judgments. The content of a moral judgment is not to be understood as involving a *proposition* that predicates a moral *property* as it is on standard Factualist accounts. Instead, it is to be seen as involving a *state of mind* that expresses a commitment to a particular *plan* or *course of action*. While predications function to describe or represent, plans function to express a motivation to act in various ways.

This semantic claim is taken to be a variety of Factualism because it is paired with an ontological claim that draws on a broader conception of the sorts of things

²⁷ It is worth noting that the rejection of Determinism for some more moderate position is plausible independently of a commitment to sophisticated expressivism. For one, we have metaphors and idioms (e.g. Quine's 'sakes'). But "literal" judgments illustrate this point too—judging that Beethoven's fifth symphony has four movements needn't entail an ontological commitment to symphonic movements (Horgan 1986).

²⁸ Calling the move to replace Determinism with Regulation the *expressivists'* insight is somewhat misleading. As we will see in chaps. 3-4, advocates of other varieties of Factualism can be understood as advocating a similar move, though to a less extreme conclusion (e.g., the non-reductive naturalism found in Boyd 1988 or Brink 1989; the response-dependent reductive naturalism of Railton 1986a or Lewis 1989). At least one of the benefits of focusing on the expressivists' rendering is that it allows us to identify (one of) the outer boundaries of Factualism.

moral facts can be. The expressivists' ontological claim is that moral facts are to be understood in a deflationary manner. More specifically, their deflationism about moral facts (like deflationism generally) maintains that there are moral facts, but insists that in granting this, they are not granting that moral facts are metaphysically substantive things that the realists and constructivists posit. Roughly speaking, to be a fact requires that there be a contentful mental state, but *not* that the content be descriptive.

At this point, the details of the expressivists' deflationism need not concern us.²⁹ What matters is that we see their insight as allowing us to distinguish between two types of Factualism—call them Descriptive and Non-Descriptive.

Descriptive Factualism: Moral judgments are objective in that they express contentful states—e.g., propositions—that describe or represent a distinctive realm of moral facts in virtue of predicating moral properties and relations to acts and individuals.

Non-Descriptive Factualism (or Expressivism): Moral judgments are objective in that they express contentful states—i.e., non-propositional commitments or plans. These states do not describe but rather evince how one plans to act under a particular set of circumstances. In this way, moral judgments circumscribe (but do not refer to) a distinctive realm of moral facts.³⁰

These refinements allow us to revise our initial conception of logical space. Though the borders between these general families will likely be vague, we can locate some familiar meta-ethical accounts:

²⁹ Though I won't pursue the matter here, there is reason to question the expressivists' move to distinguish between "robust" and "deflated" metaphysical commitments. Questions the expressivist would need to address include: What is a metaphysically deflated commitment? How does it differ from a robust one? What justifies making the distinction with regard to moral discourse?

³⁰ For more on the expressivists' rejection of Descriptive Factualism for Non-Descriptive Factualism see, for instance, Blackburn 1984: 197-210, Hogan & Timmons 2006b, and—in a more cautious form—Gibbard 2003: 182-4.

Table 2: A Revised Characterization

Semantic Function	Ontological Structure	Metaphysically Real?	
		Yes	No
Factual, descriptive	Independent	Analytic naturalism, ³¹ non-reductive naturalism, ³² non-naturalism, ³³ divine command theories, ³⁴ ...	Error theories
Factual, descriptive	Constructive	Practical reason theories, ³⁵ conventionalism, ³⁶ response-dependence theories, ³⁷ ...	Error theories
Factual, non-descriptive	Deflationary	Sophisticated expressivism ³⁸	Error theories ³⁹
Non-Factual	N/a	N/a	Emotivism, prescriptivism

We can also systematize the relations between these views as follows:

³¹ E.g., Jackson 1998.

³² E.g., Boyd 1988, Brink 1989, Sayer-McCord 1988, Sturgeon 1985b.

³³ E.g., Moore 1903, Shafer-Landau 2003, Wedgwood 2007.

³⁴ E.g., Adams 1973, 1987.

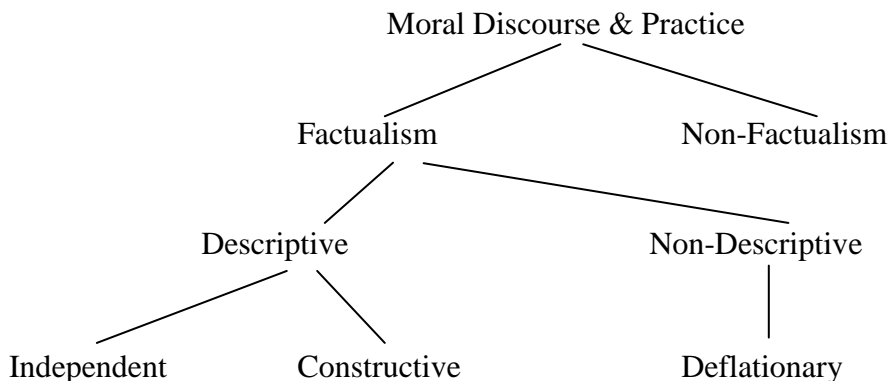
³⁵ E.g., Darwall 2006, Gauthier 1985, Korsgaard 1996, 2009, Scanlon 1982, 1998.

³⁶ E.g., Harman 1975, 1977, Wong 1984, 2006.

³⁷ E.g., McDowell 1985, Railton 1986a,b, Smith 1995, Wiggins 1987.

³⁸ E.g., Blackburn 1984, 1993, 1998, Gibbard 1990, 2003, Horgan & Timmons 2006.

³⁹ Though I will not discuss the view here, one might advocate an error theoretic version of Deflationary Factualism by, for instance, endorsing an error theory about the psychological notions (e.g., conative states like desires) on which it depends (c.f., Churchland 1981).

Diagram 1: A Map of Logical Space

4. What Success Would Look Like

If we are to develop a non-skeptical account of the objectivity of ethics, we must establish what the success of this project requires. We can begin by leveraging a move implicit in the adoption of Phenomenological Regulation—namely, identifying the semantic and non-semantic “marks of objectivity” that appear to underlie our moral discourse and practice. With these marks in hand, we will have the tools to more precisely distinguish between skeptical and non-skeptical accounts, and to specify what providing a successful, non-skeptical account would require.

(4.1) The Marks. Virtually all meta-ethical accounts grant that our thought and talk about moral matters seems to presuppose certain marks of objectivity.⁴⁰ Very generally, we can distinguish the following two types of objective marks: those that

⁴⁰ See, for instance, Blackburn 1984, 1998, Boyd 1988, Brink 1989, Darwall, Gibbard, & Railton 1992, Gibbard 2003, Horgan & Timmons 2006, Mackie 1977, Miller 2003, Railton 1986a, 1989, Smith 1995, Sturgeon 1986, and Timmons 1999.

concern the semantic and logical features of our moral discourse, and those related to our epistemological and critical practices.

Semantic and Logical Features. As we have seen (§2.1), moral judgments have characteristically descriptive semantic and logical features that strongly suggest that their underlying function is that of making claims about an objective realm of moral facts: Moral judgments typically have declarative surface syntax and are used in indicative mood. As such, they seem to be making assertions, many of which we take to be true. Moral judgments can also be embedded within truth-functional constructions, propositional attitudes, and ‘that’ clauses. They figure in a variety of legitimate and illegitimate inferences—be they deductive or inductive. Furthermore, we typically take moral judgments to have universal scope: A judgment that x is wrong in conditions C applies not just to me, but to anyone in C .

Features of Our Epistemological and Critical Practices. As we have seen, moral judgments are the product of thought and deliberation—they are a result of employing reasons (moral and non-moral) for and against various moral considerations. Furthermore, in our deliberations, we appear to attribute a distinctive weightiness to moral judgments—they are more significant with regard to our thinking and acting than our tastes, preferences, and personal ideals. In fact, that we have such deliberative practices implies fallibility, and so some sort of objectivity. This implication gains further support from other critical practices that we find in our moral discourse. For instance, we take there to be genuine moral disagreement (i.e., disagreeing parties are disagreeing about whether some moral judgments is, in fact,

true). Additionally, we presume that we could be wrong about a particular moral judgment, and that our ability to make sound moral judgments could be better. We also take ourselves to have moral knowledge and to be capable of becoming more knowledgeable about moral matters. All this suggests that our moral discourse and practice is about an objective realm of moral properties and facts.

(4.2) The Status of the Marks. Though most philosophers identify the above marks as characteristic of our moral discourse and practice, there is little agreement as to their metaphysical and epistemological status. In fact, it is often the case that *nothing* is said on this point. Those who acknowledge the issue typically fall into one of three camps. First there are those who only give the question cursory attention. They chalk the marks up as “platitudes” or “commonsense” claims or intuitions, but then say nothing about what this means.⁴¹ The trouble here is that since this type of characterization gives no substantive answer about the status of the marks, we’re left without any understanding of either how they could serve to constrain an investigation into the objectivity of ethics, or why we would be justified in claiming that they do. A general appeal to epistemic conservatism might give some legitimacy, but it would be quickly exhausted in the face of skepticism about the objectivity of ethics. Worse, these accounts appear incapable of providing a principled explanation for how to adjudicate conflicts between considerations of epistemic conservatism, and considerations from other general theoretical virtues like simplicity.

⁴¹ E.g., Mackie 1977.

The second camp falls to the other extreme in the sense that they make implausibly strong claims about the metaphysical and epistemological status of the marks—i.e., that they are *a priori* necessities.⁴² But here there are several problems. For one, barring an appeal to an analytic/synthetic distinction, it is hard to see how one could substantiate the claim that we know *a priori* that truth-aptness (say) is an essential feature of moral judgments. Furthermore, to be plausible, an appeal to an a/s distinction would have to be principled—something that is notoriously hard to do. Second, even were one able to specify some of the marks as *a priori* necessities (e.g., by say following Hare (1952) and others with regard to the distinctive reasons responsiveness of moral judgments), others—like the embeddability of moral judgments in logical connectives—seem to be clear cases of empirically discovered contingent facts about how moral language works.⁴³

The third option takes the marks to be empirical data: They are claims or hypotheses about the logical, semantic, metaphysical, epistemological, and psychological foundations of our moral discourse and practice.⁴⁴ They are what our moral thought and talk appear to presuppose in the sense that they are claims about what would have to be the case in order for our moral discourse and practice to function as it does. But importantly, this third option leaves open—at least initially—what the answers to substantive questions about the metaphysical status of these marks. For instance, it leaves open whether or not a particular mark is in fact essential

⁴² E.g., Smith 1995, Jackson 1998, Blackburn 1998, Railton 1998: 147n23.

⁴³ Blackburn's discussion of the language E_{ex} makes this point nicely (1984: 193-6).

⁴⁴ E.g., Darwall, Gibbard, & Railton 1992, Railton 1989, Timmons 1999.

to our ability to think about, talk about, and respond to moral questions. Because this option does not suffer from the defects of the other two views, and because it offers us greater flexibility in how we might ultimately treat the objective marks, it is the account that I endorse.

(4.3) From Marks to Success Conditions. Given the wide consensus about these marks, and given their status as empirical data about our moral discourse and practice, we can formulate some conclusions about what providing a successful vindication of our objectivity talk requires. First, we can sharpen the distinction between skeptical and non-skeptical accounts. Skeptics maintain that some of the objective marks cannot or should not be accommodated. Non-skeptical accounts deny this—some variety of full accommodation is possible.⁴⁵ Second, our discussion thus far (especially §2.4) strongly suggests that we should favor meta-ethical accounts that allow us to accommodate all of the objective marks. This makes accommodating all the marks a necessary (but likely insufficient) condition on any successful, non-skeptical account of objectivity.⁴⁶ Third, our discussion also suggests that it is unlikely to possible to provide an account that takes each mark at face value.⁴⁷ The result is that one must seek to privilege certain objective marks as requiring full-

⁴⁵ I don't mean to imply that there is a sharp line between skeptical and non-skeptical accounts. But I think it is rather clear that some views will be more skeptical, others less.

⁴⁶ In the chapters that follow, I will discuss ways in which merely accommodating all the marks proves to be insufficient for a successful, non-skeptical account of the objectivity of ethics.

⁴⁷ Darwall, Gibbard, & Railton (1992: 185-6) take this point to be one of the principle upshots of their review on contemporary meta-ethical inquiry.

fledged accommodation, and then arguing that some (potentially deflated) notion of the rest can be provided.⁴⁸

Finally, and most importantly, we can specify a strategy for investigating the objectivity of ethics. Rather than beginning with an antecedent conception of what the right way to distinguish the marks that should be prioritized from those that should be deflated, we should adopt a more methodologically neutral method of investigation. For instance, we should not start with a metaphysically loaded notion of (say) moral improvement (e.g., improvement involves the convergence on a mind-independent reality, or improvement only concerns increases in the coherence of one's moral beliefs). Rather, we should begin with a metaphysically neutral characterization of the data. For instance, rather than requiring that some substantive account of moral improvement be met, we demand an account that accommodates the observation that we want to become, and feel we can become, better moral judges. Success then would involve providing an account that successfully accommodates the neutrally characterized marks.

5. The Plan of Attack

My argument proceeds as follows. In chapters 2-4, I develop and apply two general tests. The first test reflects the semantic and logical marks of objectivity discussed above. In applying it, we learn about the semantic dimension of objectivity

⁴⁸ Because it will often be unclear as to what is a deflationary accommodation and what is a skeptical conclusion, the line between skeptical and non-skeptical strategies will be somewhat vague. Railton (1989, esp. 158-9) makes a similar point.

(§1.1): morality is best understood on a Descriptivist, not an expressivist, model. The second test focuses on the critical and epistemological practices underlying our moral discourse. Applying this test helps us come to a better understanding of the ontological dimension of morality (§1.2). Here the core lesson here is that vindicating the objectivity of ethics requires a metaphysically and epistemologically modest account of moral facts. Chapters 5-6 then develop and defend a distinctive alternative that is shaped and informed by insights culled from the earlier chapters. I develop a novel constructivist account that, unlike existing proposals, is able to meet our two tests. Moreover, the resulting proposal is well positioned to vindicate the objectivity of ethics because it explains moral facts in terms of familiar features of our social and psychological lives.

Chapter 2: Logic from Morals, Morals from Logic

From the discussion of chapter 1, we know that vindicating the objectivity of ethics requires demonstrating that some form of Factualism is true. Factualism, recall, is the claim that moral judgments are objective in that they express contentful states that seek—in some way—to make claims about moral facts. So Factualism is false only if moral judgments do not express contentful states—that is, Factualism is false only if a skeptical meta-ethical view—non-cognitivism—is true. But we have also seen (chapter 1.3) that Factualism comes in two varieties—descriptivism and expressivism. While descriptivists take moral judgments to be in the business of describing or making representational assertions about the world, the expressivists denies this. Surface syntax to the contrary, expressivists maintain that moral judgments express motivationally laden states of mind—in this way, they are like a command or a cry of pain.

The fact that expressivism denies that moral judgments function descriptively has significant implications for what they say about language and logic. For notice, if moral judgments don't describe, we can't appeal to standard, truth-conditional theories to give an account of either the meaning of moral sentences, or the inferential relations between sentences. Expressivists readily acknowledge that the plausibility of their proposal turns on their ability to provide a viable, alternative account of meaning and

inference.¹ Recognizing this points to a way to test whether expressivist varieties of Factualism could provide a viable foundation for an Objectivist account of ethics.

In order to give substance to this proposal, I will focus on the intuitive distinction between the logical and the extra-logical. To get a handle on this distinction, let's compare a logical inference like

(L) Nancy acted rightly; so, someone acted rightly,

and an extra-logical inference like

(EL) Murder is wrong; so, hiring a hitman is wrong.

Intuitively, both (L) and (EL) are legitimate inferences. But only (L) is logically valid in the sense that the legitimacy of (L) turns on only the facts of logic while the legitimacy of (EL) turns on, among other things, the facts of morality.

The need to distinguish between logical and extra-logical varieties of inference, entailment, validity, and consistency has played a prominent role in meta-ethical debates between expressivists and descriptivists. These intuitive distinctions are not of mere theoretical interest: We take logical inferences to have a privileged place in thought and inquiry; we see logical errors as involving a special type of failing and as inviting a distinctive sanction. So it's important that we be able to explain what distinguishes these logical notions from their extra-logical counterparts. The standard explanation is familiar: Logical consequences, for instance, are truth preserving in the sense that—no matter what interpretations we give to the extra-

¹ See, for instance, Blackburn 1984: 189-96, 1988, 1998: 72, Gibbard 1990: chap. 5, 2003: chaps. 3-4, and Horgan & Timmons 2006: 277-82.

logical vocabulary—they will not lead from true premises to false conclusions. But clearly this standard explanation is not available to expressivists, they must provide an alternative to the standard, truth-centric explanation. However, the plausibility of the proposed alternatives is deeply contested.²

While this traditional debate has drawn out important and interesting challenges to the expressivists' program, it overlooks a deeper and more significant issue. For notice, not only is the standard explanation of logical consequence framed in terms of the preservation of truth, but it also assumes that we can distinguish the logical terms from the extra-logical ones. That is, it presumes we have an account of *logical form*. To date, questions of logical form have not been part of debates between expressivists and descriptivists.³ That's a mistake. For one, questions of logical form are, in an important sense, prior to questions concerning the preservation of truth: Logical consequences are truth-preserving *in virtue of* their logical form; it's logical form that *guarantees* that true premises won't lead to a false conclusion.⁴ Therefore, any complete account of logical consequence—truth-centric or expressivist—must include an account of logical form that explains why logical consequences carry this guarantee. Moreover, we shouldn't just assume that a viable account of logical form

² Since the problems with the expressivists' alternative accounts of logic and inference (often lumped together under the label "the Frege-Geach problem") are familiar, I will merely mention some of the more prominent ones—expressivism cannot provide a viable account of negation, it can't make sense of normative content in embedded contexts, it conflates failures of logical reasoning with more general failures of rationality. For a sampling of the representative arguments, see Hale 1993, Schueler 1988, Wright 1988, van Roojen 1996, and Unwin 2001.

³ This may be too strong: the arguments of Hale 1993 and van Roojen 1996 seem to raise concerns in the vicinity of this second feature of logical consequence.

⁴ On this point, see Tarski 1935, Etchemendy 1990, Sainsbury 1991, Sher 1991, and MacFarlane 2005 among others.

can be given, for it is controversial whether descriptivists—much less expressivists—can provide a principled account of logical form.⁵

But once we recognize the importance that matters of logical form have for our understanding of logical consequence, we see that it will likely have implications for debates between expressivists and descriptivists. For one, it points to a new way of assessing the adequacy of the expressivists' alternative account of logical consequence: in addition to asking if their proposals are truth-preserving, we can also ask if they are grounded in a principled account of logical form. Yet recognizing the importance of logical form also has implications for descriptivists. In light of the above, we can now see that the traditional descriptivist arguments against expressivist accounts of logic have force *only if* descriptivists themselves can provide a principled account of logical form.⁶

In what follows, I will argue that descriptivists are better positioned than their expressivist rivals to provide the needed account of logical form, and so better able to capture the intuitive distinction between logical consequences and extra-logical ones. This conclusion is significant for several reasons: First, it provides a new argument against expressivism—one that demonstrates that even recent, sophisticated proposals

⁵ Tarski, in his seminal paper on logical consequence, both noted that his account relied crucially on a division of logical and extra-logical vocabulary, and expressed unease about the possibility of a principled demarcation (1935: 418-9; c.f., 1986). Others—e.g., Etchemendy 1990—are more thoroughly skeptical. Also, though most logicians endorse a truth-centric account of the meanings of logical terms, there are some (e.g., Brandom 2001) that pursue “expressivist” alternatives.

⁶ See, for instance, the citations in note 3.

for an expressivist-friendly account of logical consequence fall short.⁷ Second, we will see that descriptivists can make use of this new argument (and also substantiate their traditional objections) only if they are willing to take a controversial—but, I will argue, plausible—stand on claims about the nature and foundations of logic. The overall upshot is that vindicating the objectivity of ethics requires a descriptivist variety of Factualism, not an expressivist one.

1. Background: Logical Consequence and Logical Form

To understand why expressivists face a special problem in making sense of logical consequence, we need a better understanding of what it means to say that logical consequences are determined by logical form.⁸ As we will see, the standard explanation holds that logical consequences are valid solely in virtue of the meanings of their logical terms: logical terms specify the underlying form of logical consequences, the having of which *guarantees* that their conclusions follow from their premises. A closer look at this explanation of the relationship between logical form and logical consequence will allow us to formulate a general and intuitive constraint—a modal test—that any viable account of logical consequence must meet.

(1.1) The Importance of Logical Form. Philosophers of logic note two distinctive features of logical consequences. Most obviously, logical consequence relations are *formal* in the sense that they are determined solely by the structure or

⁷ E.g., Horgan & Timmons (2006: 288-97), and Schroeder (2008: chap. 5).

⁸ For simplicity, I will focus for the most part on logical consequence—though what I say applies with equal force to related logical notions.

form—not the content—of the sentences involved: Intuitively, it's in virtue of *form* alone that the conclusion 'someone drank the beer' follows from the premise 'Bob drank the beer.' By contrast, extra-logical consequence relations are not purely formal—the validity of extra-logical inferences relies on the substantive *content* of the sentences involved: the conclusion 'beer is made of water' follows from premise 'beer is made of H₂O' because of what terms like 'water' and 'H₂O' mean. Moreover, which terms we identify as contributing to the form (rather than content) of a sentence matters: The terms we identify as logical—and so formal—determine which inferences we count as logical consequences. For instance, we take 'someone' to be a logical term. That we do this helps explain why

(1) someone is loud and rowdy

is a logical consequence of

(2) Bob is loud and Bob is rowdy.

But it also explains why (1) is not a logical consequence of

(3) someone is loud and someone is rowdy.

This is significant: the terms that we take to be constitutive of logical form determine what counts as a logical consequence; this means that we need a principled way of identifying which terms are the logical terms.

Enter necessity. Philosophers of logic take the second distinctive feature of logical consequences to be that they are *guaranteed* to be truth preserving. Witness John Etchemendy:

The most important feature of logical consequence, as we ordinarily understand it, is a modal relation that holds between implying sentences

and sentences implied. ... [The] conclusions [of logically valid arguments] are said to “follow necessarily” from their premises. (1990: 81)

Given the significance of this guarantee to our understanding of logical consequence, philosophers of logic have traditionally taken the correct selection of logical terms—and so the correct specification of logical form—to be the one that captures this distinctive necessity.⁹

(1.2) The Modal Test. Combining these two features—formality and necessity—points to a general constraint on any plausible account of logical form: A viable account must specify a set of logical terms that captures the guarantee we associate with logical consequences. But if we’re to make use of this constraint, we need to know more about the guarantee. Intuitively, logical consequences are strongly counterfactually stable: they continue to hold in a world w even in the face of alterations to what is (say) nomologically or metaphysically possible in w . By contrast, extra-logical consequence relations exhibit lesser degrees of counterfactual stability. This conclusion is well supported by our modal intuitions.¹⁰ For instance, we take logic to be *more basic* than things like science, psychology, and metaphysics in the following sense: what’s logically possible does not seem to depend on things like what is scientifically, psychologically, or even metaphysically possible. Moreover, a

⁹ On the nature of the formality and necessity of logical consequence, see for instance Tarski 1935, Sher 1991: chap. 3, Etchemendy 1990, and Sainsbury 1991: chaps. 1, 6.

¹⁰ Tarski emphasized the importance of our ordinary intuitions for philosophical accounts of logical consequence: “Certain considerations of an intuitive nature will form our starting-point” (1935: 414). Also see Sher 1991: chap. 3 and Etchemendy 1990: chap. 6.

converse dependency does not hold.¹¹ We also take worlds where the facts of logic are different—e.g., worlds where disjunctive syllogism is not a valid inference schema—to be *more remote* (if there even are such worlds) than worlds where the extra-logical facts are different—e.g., worlds where objects can move faster than the speed of light, or where reddish-green things are just as common as bluish-green ones. In fact, we can explain why we have these intuitions in a way that deepens our understanding of why logical consequence has the distinctive modal profile that it does: Logic is both our most general mode of inquiry, and is “topic-neutral” in the sense that it does not depend on facts about things like science, psychology, and (much of) metaphysics. It is because of these features—generality and neutrality—that logical consequence has its distinctly stable modal profile.

Given this account of the modal features of logical consequence, we can formulate a substantive constraint on any viable theory of logic form:

The Modal Test: A viable theory must explain what it is about logical form that provides logical consequences with their distinctive modal stability.

With this test in hand, we’re positioned to see both why expressivists have difficulties providing an account of logical consequence, and how descriptivists can avoid them.

2. Meta-Ethics and the Modal Test

At this point, one might think that meeting the modal test is a problem in the philosophy of logic, not a problem in meta-ethics. But that would be a mistake.

¹¹ The claim in the text is meant to be restricted to the non-logical aspects of metaphysics.

Though they may not realize it, both expressivists and descriptivists are—by their own lights—committed to meeting the modal test. As we’ve noted, expressivists acknowledge that the plausibility of their accounts requires that they be able to capture the “mighty ‘musts’ of logic.”¹² But, as the discussion so far reveals, the characteristic feature of these “musts” lies in the distinctive guarantee that they carry—their uniquely stable modal profile. This means that capturing what’s special about (say) the “musts” of logical consequences will require expressivists to meet the modal test. But descriptivists face the same requirement. The importance of this point has not been appreciated: the traditional descriptivist critiques of expressivist accounts of logic presume that a viable, descriptivist response to the modal test can be provided. But that’s an assumption that we—and especially the expressivists—should challenge.¹³ In light of all this, it is surprising that existing debates between descriptivists and expressivists on matters of logic and inference have overlooked the importance of questions of logical form.¹⁴

At this point one might worry that the modal test places an excessive demand on expressivists: given that expressivists acknowledge that they’re providing an *alternative* account of logic and inference, demanding that they follow the *standard*

¹² The quote is from Blackburn (1998: 72), but also see his (1984: 189-96) and (1988) as well as Gibbard (1990: chap. 5, 2003: chaps. 3-4) and Horgan & Timmons (2006: 277-82).

¹³ It’s worth emphasizing that meeting the modal test is not a trivial task. There is significant debate among philosophers of logic regarding how—even whether—it can be done. For a representative sample, see Tarski 1935: 418-10; 1986, Etchemendy 1990, Dummett 1991: 54-5, Sainsbury 1991, Sher 1991, and MacFarlane 2005.

¹⁴ The assumption that there is a principled way of distinguishing between logical and extra-logical content is explicit in the recent expressivist accounts of logical consequence proposed by Horgan & Timmons (2006: 288-97) and Schroeder (2008: 69-70), and comes out in the work of Blackburn and Gibbard via their use of inference schemas. It is implicit in most descriptivist critiques.

procedure of grounding an account of logical consequence in an account of logical form just asks too much.¹⁵ This concern would be legitimate were the modal test to require a truth-centric account of logical form. But it doesn't. It merely requires an account of logical form—truth-centric or otherwise—that captures the distinctive modal stability that we associate with logical consequences. Granted, because of the expressivists' self-imposed prohibitions on the use of robust notions of truth, satisfaction, and the like, they have fewer resources from which to explain this stability (more on this below). But the conclusion that they cannot develop a viable account isn't built into the test itself. Moreover, as we will see, there is a substantial tradition in the philosophy of logic that maintains—in line with the expressivists' foundational commitments—that the necessity of logic is grounded in basic features of thought.

In sum, the modal test provides us with a new way of assessing debates between expressivists and descriptivists. Whereas the traditional debates focused on whether expressivists could provide an account of logic that was truth-preserving, the modal test asks both expressivists and descriptivists to explain what philosophers of logic take to be the most significant feature of logical (as opposed to extra-logical) consequences—namely, their distinctive modal stability.

¹⁵ Thanks to an anonymous referee for helping me see the need to explain why it's reasonable to demand that expressivists meet the modal test.

3. The Limits of Expressivism

To assess expressivists' ability to meet the modal test, we must say a little more about what expressivism is. First, expressivism is, at its core, a thesis about the nature of thought. It holds that normative claims do not aim to describe the world, but rather express distinctive non-cognitive states of mind. So, in judging (say) that torture is wrong, one does not make an assertion about the way the world is; rather, one's speech-act is more like a cry of pain or a command to close the door. But since logic and reasoning are normative, the expressivist is committed to treating them in the same way.¹⁶ That is, they are committed to grounding an account of logic, not in some robust notion of truth or satisfaction, but rather in claims about the nature of certain mental states.¹⁷ Second, expressivists have a standard strategy for addressing charges that they cannot do something that their descriptivist rivals can: They determine how the problem is avoided within the descriptivist framework, and then provide an expressivist model that mimics it.¹⁸ So while philosophers of logic typically express their accounts of logical consequence and logical form in ways that make essential use

¹⁶ A couple of points: (1) The normativity of logic and reasoning is not the only reason expressivists are committed to treating them non-descriptively. There is also, for instance, the problem of fitting non-truth-conditional normative sentences within the standard, truth-conditional accounts of logic. (2) Hybrid views—views that seek to combine an expressivist account moral content with a traditional, truth-centric account of logic and inference—would reject the claim in the text that expressivism about (say) morality entails expressivism about logic. But they are not the concern of this essay. See Schroeder 2009 for an insightful discussion of different forms of hybrid views and a (largely) skeptical assessment of their viability.

¹⁷ See, for instance, Blackburn 1984: 191-5, 1988: 190-3, Gibbard 1990: 98-102; 2003: 62-3, Horgan & Timmons 2006: 277-8, and Schroeder 2008: chaps. 3-5.

¹⁸ See Gibbard 2003: 81 for a nice statement of this standard move.

of substantive notions like truth and satisfaction, we can use this strategy to develop expressivist-friendly versions that mimic these accounts.

There are two very general strategies that philosophers of logic have pursued in order to explain the characteristic guarantee of logical consequence. The first argues that logical form provides the underlying structure of thought and language: it's because logical form captures certain special features of thought and language that logical consequences must preserve truth. The second maintains that logical form is an abstract, extra-mental part of the world—it's in virtue of the fact that logical terms map to distinctive formal features of the world that logical consequences carry the guarantee that they do.¹⁹ Since strategies that look to thought and language are more in line with the expressivists' framework, let's start by taking a closer look at them.

(3.1) Appeals to the Nature of Thought and Language: Version 1. One way to flesh out the details of the first strategy is by following Davidson and Quine.²⁰ Proposals of this sort take logical form to play an essential, structural role in a systematic account of the meaning of a given language. In particular, logical form is a way of regimenting the truth-functional relations of the sentences of a language so that the (basic) inferences that users of that language make are represented as truth preserving. To flesh this out, consider the following explanation of why I took the cab to the airport rather than the bus: I believed that I could only get to the airport by cab

¹⁹ Though this is not the standard way of parsing accounts of logical form (but see Sher 1999), I use it because it proves to be a particularly helpful way to explore the modal test with the context of expressivism and descriptivism.

²⁰ See, for instance Davidson 1984; Quine 1960, 1980.

or by bus; I believed that a bus would not get me there as quickly as a cab; and I wanted to get to the airport as soon as possible. But—importantly—this account gets its explanatory power by representing my decision process as having the form of something like disjunctive syllogism. Furthermore, in order for this account to explain my decision to take the cab, two conditions must hold: (i) I must be understood as taking terms like ‘or’ and ‘not’ to have their standard meanings, and (ii) these logical terms have their standard meanings *because* that’s how one represents my reasoning as truth-preserving.

This fact—namely, that the above two conditions must hold in order for the above account to explain my decision to take the cab—could be used to explain the relationship between logical form and the distinctive necessity of logic in a way that meets the modal test: Logical form (as specified by the standard meaning postulates of ‘or’, ‘not’, and the like) articulates the formal features of thought that are essential to our capacity to think, speak, and act as we do. In short, this first proposal maintains that it is because these formal features *must* be in place if our reasoning is to be truth preserving that logical consequences have their distinctive necessity.²¹

In its canonical form, this strategy for meeting the modal test is built around robust semantic notions of truth and satisfaction. However, there is an expressivist-

²¹ Notice that the argument in the text does not establish that the standard selection of logical terms is the *only* way to explain our reasoning as truth-preserving, and so does not (yet) explain why the formal features of language and thought *must* be as they are. Clearly an expressivist (or descriptivist) interested in pursuing this strategy would need to demonstrate that the standard selection of logical terms is the only way to secure a satisfactory explanation of our reasoning. In what follows, I will assume that the needed explanation can be provided.

friendly alternative in the vicinity: rather than taking the correct account of logical form to be captured by what's necessary for the preservation of truth, the expressivist can take it to be that which is essential to our ability to think and act as the rational beings that we are. That is, rather than grounding the account in a substantive notion of *truth*, the expressivist relies on an account of *rationality*.²² More specifically, the idea is this: We are (generally) rational beings in the sense that our behavior is shaped and constrained by the systematic connections that hold between our beliefs, desires, and intentions. We also want a way of explaining, for instance, why I took the cab that—given my beliefs, desires, and intentions—takes my behavior to be rational. Getting such an account requires, among other things, that I be understood as taking ‘or,’ ‘not,’ and the like as having their standard meanings. So, as with the canonical versions of this strategy, the expressivist-friendly version takes logical form (as specified by the standard meaning postulates of ‘or,’ ‘not,’ etc.) to articulate the formal features of thought that are essential to our capacity to think, speak, and act as the rational beings that we are.

To see why the account of the distinctive necessity of logic that we get from this proposal fails to meet the modal test, let's begin by considering what it says about the nature of logical consequence. We can do this by returning to the cab vs. bus

²² Blackburn (1998: 51-9, 68-75) and Gibbard (1990: 99-102, 156-70) make use of roughly this line of reasoning to give an account of extra-logical content. It is unclear whether they would endorse the move in the text to apply the strategy to flesh-out their accounts of logical content as well. Regardless, they and Horgan & Timmons (2006: 278) have the resources to make a version of the argument that follows. However, an expressivist inclined toward this strategy needs to provide an account of rationality that is not cashed out in terms of truth-satisfaction or property exemplification. Both Blackburn 1998 and Gibbard 1990 make proposals. While I will not assess these here, see Sturgeon 1995 for concerns.

example. This example draws out that what counts as being rational is a function of, among other things, the nature of our psychology; it's a function of the nature of the connections between our beliefs, desires, intentions, etc. But surely our psychological constitution could have been different. For instance, we could have been less responsive to psychological phenomenon like cognitive dissonance.²³ If this were the case, we would be less "disturbed" when we found ourselves to be in mental states where we affirmed sets of judgments like $\{P \text{ or } Q, \text{ not-}P, \text{ not-}Q\}$, and so more likely to find ourselves in such states. But if we were largely unaffected by such dissonance, understanding our behavior as rational might require denying that logical terms like 'or' and 'not' map onto the standard rules of logic (i.e., the \vee and \sim introduction/elimination rules); it might require denying that disjunctive syllogism counts as a legitimate formal inference.²⁴

The implications of this possibility are significant. For one, because this proposal must ground the modal profile of the standard rules of logic in substantive claims about how human psychology happens to be, it makes the modal profile of these rules (and, therefore, the modal profile of logical consequence) a function of the modal profile of our psychology. So it commits the expressivist to the dubious claim

²³ This is not implausible given the varying degrees of cognitive dissonance that social psychologists have found in their subjects. See, for instance, Aronson 1999.

²⁴ Given the expressivists' commitment to naturalism, I see no way for them to deny this possibility. But presumably we're dealing with a conceptual question here, and so can set questions of naturalism aside. Doing this allows us to consider more fanciful cases: Suppose that it is well known that an evil demon will put a malicious curse on us if we reason by disjunctive elimination. Under such circumstances, understanding our behavior as rational might again require denying that 'or' and 'not' have their standard meanings.

that the modal stability of human psychology is on par with the modal stability of logical consequence. For similar reasons, the proposal also makes logical consequence dependent on contingent features of human psychology. These results entail that this proposal fails to meet the modal test—it cannot explain why the modal profile of logical consequence is distinct from, and more stable than, the modal profile of human psychology.²⁵

(3.2) Appeals to the Nature of Thought and Language: Version 2. If an appeal to a general interpretive constraint proves too weak to secure the distinctive necessity associated with logical consequence, expressivists might try to secure a stronger foundation by looking to versions of the first general strategy that take logical form to be the upshot of certain privileged rules of thought and reasoning. Philosophers of logic who have pursued this proposal identify the privileged set of rules with the most basic rules of inference—rules that are universally applicable to judgment and inference.²⁶ Furthermore, it seems that something like this idea can be found in Gibbard's (2003) proposal.²⁷ Taking a closer look will help us see how an expressivist might develop it into a response to the modal test.

²⁵ The expressivist strategy of grounding logic in an account of rationality faces further problems. Mark van Roojen (1996) argues that it results in expressivists conflating logical consequences with rational consequences. Though van Roojen's argument is controversial (see, for instance, Gibbard 2003: 75-9), the problem he identifies is on point: our notion of what's rational tracks something different than does our notion of what a logical consequence is. In fact, the modal test allows us to sharpen van Roojen's insight: we can now specify *why* a move to ground logic in rationality must fail—it lacks the necessary modal profile.

²⁶ A diverse group follows strategies of this general form—for instance, Frege 1893/1903, Hacking 1975, Dummett 1991, and (in more epistemological form) Peacocke 1976.

²⁷ Blackburn's discussion of decisions and decided states has a similar structure (1998: 68-74). In fact, he presents it by drawing on an earlier, unpublished version of Gibbard's (2003) proposal.

Gibbard focuses on the mental operations we employ when making practical decisions about what to do. The result of these operations is a complex, non-cognitive mental state—something like a set of inter-related commitments. Following Gibbard, let's call this basic mental operation planning and the resulting mental state a plan. For the purposes at hand, plans and planning are interesting for several reasons. First, plans seem to involve certain formal features (or operations) essentially. For instance, I can *combine* parts of plans—I can combine my plan to go to the grocery with my plan to pick the kids up from school; I can *reject* plans—I can reject a plan for a 60th birthday party in my honor; I can make *general* plans—I can plan to vote for whichever Democrat wins the nomination.²⁸ Second, the mental operations of planning that are essential features of plans appear to match up nicely with the formal features of logic. Witness Gibbard:

Combining, rejecting, and generalizing apply to mental operations [of planning], which then have content. The content is expressible with the logical operators of conjunction, negation, and quantification. These logical devices mirror the mental operations of combining, rejecting, and generalizing. ... Logic ... is a matter of the ways statements allow determinate possibilities and rule them out. The logical import of a statement, we can say, is a matter of [what] it allows and [what] it rules out. This gives us a way of thinking of the content of such judgments.
(58)

Given the context, it seems that by “content” Gibbard intends to refer to something like the meaning postulates (rules) standardly associated with ‘and’, ‘not’, and ‘all’. So

²⁸ Though Gibbard (2003: 53-9) only considers combining, rejecting, and generalizing he, presumably, does not intend this to be an exhaustive list of the features that expressivists could take as essential to planning.

we could see Gibbard as suggesting that logical form merely regimentes the foundational operations that underlie the basic non-cognitive mental operation of planning. This points to a possible response to the modal test: Logical consequences have their distinctive modal stability because logical form expresses a set of mental operations (the rules of logic) that are essential to planning; it's *because* these operations are essential to planning that logical consequence has its distinctive necessity.²⁹

To evaluate this proposal, we need a better understanding of how well these mental operations map onto the standard rules of logic. So let's consider general plans in relation to the rules of logic governing 'all'.³⁰ Suppose I make a general plan to believe whatever predictions my guru makes—*any* prediction he makes, I plan to believe. We can now ask how expressivists explain this 'any'. Gibbard's move (2003: 53-9) is to appeal to a radically idealized notion of planning. On this understanding, general plans are understood to be "maximally complete" instances of planning—one has settled on a course of action for *any* situation one might be in. But, as we've noted, the expressivists' resources are restricted to appeals to features of our mental lives. So they must explain the maximal completeness of plans as a conceptual completeness—plans that prescribe a course of action for any *conceivable* situation.³¹ The upshot is

²⁹ An expressivist might seek to augment this account of what makes this set of mental operations special by mimicking the "permutation invariance" accounts of logical form found in Tarski 1986, McGee 1996, and elsewhere.

³⁰ Similar results hold for the relationship between rejecting and '~' and combining and '&'.

³¹ Gibbard (2003: 54) is quite explicit to explaining the completeness of hyperplans as a conceptual completeness. C.f., Blackburn 1998: 74n18.

that the notion of ‘any’ associated with my general plan to believe my guru’s predications will be a conceptual necessity—anything I could *conceive* of him predicting, I plan to believe.

However, since the modal features of the meaning postulates for ‘all’ are, on this account, taken to be a function of the modal features of the basic operations of forming general plans, these postulates will involve a conceptual modality. More specifically, the \forall -introduction/elimination rules turn out to be a function of our conceptual capacities in the sense that the scope of, say, \forall -introduction is constrained by the sorts of general plans that I could conceive of (e.g., the sorts of things that I can conceive of my guru predicting). This prompts two problems.³² First, to the extent that we think that the scope of ‘all’ is constrained, these constraints do not seem to be a function of the sorts of things that we can conceive of. Worse, the bounds of our conceptual capacities are a function of, among other things, our psychological endowment. But, as we’ve seen (§3.1), the move to make the modal profile of logic a function of the modal profile of our psychology does not work—it fails to get the

³² In addition to the concerns that follow in the text, it is worth noting that in wedding themselves to a Stalnaker-Lewis style possible-worlds semantics, expressivists risk inheriting additional problems associated with semantic frameworks of this sort. For instance, a possible worlds account has difficulties making sense of indexical expressions, as well as explaining why, when I believe that p , I don’t also believe everything entailed by p (see Robbins 2004 for a discussion of these and other problems). Gibbard (2003) and others (e.g., Dreier 1999) acknowledge (some of) these problems and make moves to address them. But it is not obvious the extent to which their efforts are sufficient. To the extent that they’re not, the problems with this expressivist proposal are compounded.

modal profiles associated with logical consequence right. So the appeal to the nature of planning also fails to address the modal test.³³

Let's take stock: Both of the proposals for addressing the modal test through an appeal to features of thought fall short. Though other alternatives may be available, there's good reason to think that they too must fail. Here's why: To address the modal test, expressivists must provide an account of the formal features of thought that can secure the distinctive necessity of logical consequence. But since addressing the modal test involves showing that logical consequence is more stable than other forms of consequence—psychological, rational, etc.—there is no reason to think that any account grounded in claims about the nature of thought could ever be enough. This is

³³ Objection: The argument in the texts works only if it's not possible to plan for the inconceivable. But it seems we can: one plans for everything one can conceive of, and then one plans for what to do were the inconceivable to occur. So it's wrong to think that the scope of the expressivist's notion of 'all' is too limited to meet the modal test.

Reply: The central assumption of this objection is mistaken. Given what the expressivist takes planning to be, it is *not* possible to plan for the inconceivable. There are at least two reasons for this. (1) According to the expressivist, for a mental state to be an act of planning, it must rule out possible courses of action; what makes a particular act of planning what it is (e.g., a plan to Φ and not a plan to Ψ) is that the planner has ruled out doing certain things. In light of this, consider my plan to flee to my hideout if the inconceivable occurs. When I'm planning for the inconceivable, I must have reason to believe that a particular course of action—like fleeing—will be possible (were this not the case, I couldn't take myself to be planning—i.e., ruling something out). However, it's constitutive of "planning" for the inconceivable, that one can have *no* sense of which options are possible. So I can have no reason to believe that fleeing is possible. So my "plan" to flee cannot be a real plan. (2) As expressions of practical, non-cognitive mental states, acts of planning must involve a commitment to act in a certain way: my plan to Φ expresses my commitment to Φ -ing. So for my plan to flee to my hideout to count as a real plan, it must be psychologically possible for me to commit now to fleeing were the inconceivable to occur. But notice what I'm doing. I'm deciding how to act in a situation about which I can have, by definition, no information. So I'm now committing to flee at some point in the future knowing that—when that time arrives, and I come to know the details of the situation confronting me—I'll surely have no reason to follow my plan. But there's no reason to think such a state is psychologically possible. Notice: the mental state I'd need to be in (committing to a course of action knowing that, come action time, I'll have no reason to follow through) bears a striking resemblance to Kavka's (1983) toxin puzzle (intending to drink a non-fatal toxin an hour from now knowing that, at that time, one will have no reason to fulfill one's intention). What makes Kavka's case paradoxical is that the mental state it purports to describe does not appear to be psychologically possible. This gives us a second reason to deny that it's possible to plan for the inconceivable.

significant for it means that meeting the modal test requires an *extra*-mental account of logical form.

(3.3) Appeals to an Extra-Mental Account of Formality. Philosophers of logic who take logical form to be part of the world hold that logical terms refer to abstract, formal features of the world. The appeal of this proposal is that, suitably fleshed out, these formal features of the world would have the stable modal profile needed to pass the modal test. Gila Sher offers the most worked out version of this strategy, so I will begin by sketching her account and explaining how it meets the modal test.³⁴ I will then demonstrate that an expressivist-friendly version of Sher's proposal cannot be made to work.

Sher holds that the formal features of logic are a subset of the formal properties and objects found in mathematics and set theory. So, for instance, 'and' refers to intersection, 'some' refers to non-emptiness, and 'not' refers to complement. Two features of this account are of note. First, these properties and objects are formal: their essential nature is wholly a function of their structure in the sense that they exemplify regularities or patterns the nature of which does not depend on their substantive content. For example, the property of being a set is a purely structural (and so formal) property in the sense that its essential nature does not depend on what the collected objects are (or even that there are any objects collected). It merely requires that there are (possibly empty) collections. Similarly for other set theoretic properties like non-

³⁴ See, for instance, Sher 1991, 1999. Wittgenstein 1921 offers an alternative on which, though logical form is part of the world, it is an unanalyzable feature of it.

emptiness and intersection: to have the property of non-emptiness just is to have the purely structural feature of being a set containing at least one object; to have the property of intersecting just is to have the purely structural feature of a pair of sets both of which contain some common object(s).

Second, these formal properties and objects are not mental constructions, but rather abstract features of the world; they are the formal (i.e., non-contentful) regularities that can be exemplified by particular concrete objects. Our logical terms refer to these extra-mental formal properties and objects. Sher then argues that these formal properties are related to one another in special, law-like ways. For instance, if the intersection of three sets is non-empty, the intersection of two of those sets *must* also be non-empty. Furthermore, these law-like relations are special in that they are formal: the relationships they express do not depend on the particular features of the relata, but rather on the mere fact that certain formal properties and objects will evince distinctive regularities or patterns. Because of this, the existence of the formal properties/objects guarantees that certain law-like relations will hold (e.g., the existence of three intersecting, non-empty sets guarantees that any two of those intersecting sets will also be non-empty).³⁵

³⁵ Sher (1991: chap. 3) gives a more rigorous explanation and defense of this mathematical conception of formality and its application to logic that appeals to an invariance condition—one that holds that logical constants are structures (e.g., ordered n -tuples) that are invariant under all permutations of the domain of objects. While the details of her account are interesting, for present purposes, we can stick with the bare-bones presentation given in the text. Furthermore, though others also seek to explain logical form in terms of invariance (e.g., Tarski 1986, McGee 1996), Sher's account is (to my knowledge) unique in tying invariance to an extra-mental reality.

The result is a distinctive account of the nature of logical consequence. Sher's account explains why the sentence

(4) grass is green; therefore, something is green

expresses a logical consequence by taking the logical form of this sentence to correspond to certain formal features of the world, features that are related to one another so as to guarantee that if 'grass is green' is true, 'something is green' must be as well. A more explicit explanation would go like this: To get to the logical form of (4) we begin by expressing it set-theoretically as

(4_S) the set of green things contains grass; therefore, the set of green things is non-empty.

We can now make its formal content explicit

(4_L) there is a set that contains a particular object; therefore, there is a non-empty set (symbolically, ' Ga ; therefore $\exists(x)Gx$ ').

With the formal content in hand, we can see that the logical form of this sentence appeals to a set of formal objects. Moreover, these formal objects obey a particular formal law—namely,

(F) Any set containing a particular object *must* be a non-empty set.

And so we arrive at our explanation: It's in virtue of this formal law that (4) expresses a logical consequence.

Sher's account is significant because it answers the modal test. Because her proposal identifies the logical terms and the rules of logic with formal properties and laws, logical consequences inherit the underlying modal profiles—profiles that correspond nicely with the distinctive necessity we associate with logic. More

specifically, this account gets the modal profile of logical consequence right because it takes the formal properties and objects that are constitutive of logical form to be distinctive formal features of the world; these formal properties and objects exist independently of things like the beliefs or evidence that we have about them. As abstract mathematical entities, they are “removed” from the concrete features of the world in the sense that their essential nature does not depend on what their substantive content happens to be. So things like the facts of human psychology and the laws of physics will not affect their modal profiles.

Sher’s proposal works because it identifies logical form with abstract formal features of the world. So, given what we’ve learned so far, the viability of the expressivists’ program turns on their ability to mimic something like Sher’s account. This is a substantial challenge: As we’ve seen, the expressivist takes logic and inference to be ultimately grounded in certain basic features of thought. So he needs to be able to show that *mental states* can inherit the logical properties of their formal, mind-independent content in a way that is on par with Sher’s account of how *sentences* like (4) inherit their logical properties from their formal content. In recent work, Mark Schroeder develops a sophisticated account of logical consequence on behalf of the expressivist.³⁶ However, Schroeder’s proposal presupposes that the expressivist has an account of logical form.³⁷ But for reasons that are now familiar, we can see that his account does not address the modal test and so must be seen as

³⁶ Schroeder (2008: chaps 3-5).

³⁷ See pgs. 69-70 for an especially clear statement of this assumption. A similar point holds for the recent expressivist account of logical consequence provided in Horgan & Timmons 2006.

inadequate. That said, Schroeder's proposal provides tools that will allow us to develop an expressivist-friendly version of Sher's account.

Schroeder observes that we readily accept that some mental states inherit logical properties from their contents. Consider: What's wrong with believing that grass is green and believing that grass is not green? Well, the beliefs are logically inconsistent.³⁸ Furthermore, we can explain why these *beliefs* are logically inconsistent in terms of the inconsistency of their formal *content*.³⁹ That is, it is because the formal content of the belief that grass is green (i.e., that set G contains object *a*) is inconsistent with the formal content of the belief that grass is not green (i.e., that set G does not contain object *a*) that we take the beliefs themselves to be logically inconsistent. Schroeder then notes that what works for cognitive states like belief works just as well for certain non-cognitive states like intending. For example, we explain what's wrong with intending to drink the beer, and intending not to drink the beer in terms of the (logical) inconsistency of their (formal) contents. In light of this, Schroeder concludes that it's reasonable for the expressivist to maintain that mental states like believing and intending are like sentences in that they inherit logical properties such as inconsistency from their contents.⁴⁰

³⁸ In what follows, I will for convenience, follow Schroeder in focusing on logical inconsistency rather than logical consequence.

³⁹ The discussion in the text glosses over subtle and interesting features of Schroeder's proposal. For instance, Schroeder shows how we can move from (i) a *general* notion of inconsistent belief explained in terms of *substantive* content, to (ii) an account of *logically* inconsistent beliefs explained in terms of *formal* content. For simplicity, I focus only on the later.

⁴⁰ One of the more interesting aspects of Schroeder's proposal is his demonstration that any viable expressivist account of logic must appeal to a single, general, non-cognitive mental state—something he labels "being for". (Schroeder stipulates that being for must be like belief and intention in inheriting

However, it's important to recognize that this phenomenon of inheritance is not the result of the formal content alone, but rather a result of the *combination* of the formal content and the underlying mental state. To see this, notice that—in contrast with believing or intending—there is nothing odd about *wondering* whether grass is green and *wondering* whether grass is not green. This means that only certain mental states—beliefs and intentions, but not wonderings—are what Schroeder calls “inconsistency transmitting attitudes,” attitudes that are inconsistent when and only when their contents are inconsistent. But so long as at least some non-cognitive states are inconsistency-transmitting (and intentions appear to be), the expressivist seems to have a way of showing how logical properties like inconsistency can be grounded at the level of mental states.

We're now in a position to see how expressivists could make use of Sher's proposal to address the modal test. The idea is this: Sher's account of formal properties and laws provides formal contents with right modal features, and Schroeder's account of inconsistency transmitting attitudes suggests that logical properties can be inherited by certain mental states. By combining these two ideas, the expressivist might have a way to answer the modal test. However, this strategy works only if the relevant mental states (believing, intending) can inherit the *modal* properties of their contents just like they inherit the *inconsistencies* of their contents.

logical properties like inconsistency from its contents.) Since nothing in the argument that follows turns on the details of Schroeder's account of being for, I will stick with the more familiar states of belief and intention.

That is, these states need to be what we might call “modality transmitting attitudes” in addition to being inconsistency transmitting attitudes.

But there’s little reason to think this is the case. For starters, the proposal under consideration relies on a dubious assumption. To draw it out, first recall that mental states, as psychological phenomenon, do not have particularly stable modal profiles. Yet the proposal under consideration maintains nonetheless that the formal features of the world are such that they somehow increase the modal stability of our mental states so as to put it on par with that of logical consequences. But such a remarkable feat surely requires explanation. Until we have that, there’s no reason to grant the claim of modality transmission to mental states.

But even if such an account could be provided, a further problem remains. Like with inconsistency transmitting attitudes, the expressivist needs it to be the case that only some mental states are modality transmitting. After all, if a mental state like wondering does not combine with formal content so as to inherit logical properties like inconsistency, then how could it inherit the associated modal properties? Because of this the expressivist must be able to demonstrate that the formal features of the world only increase the modal stability of *select* mental states. How might he do this? Presumably, by following his standard strategy: he will identify the way that a descriptivist like Sher explains how sentences inherit their modal stability, and then seek to mimic this solution at the level of mental states.

Since Sher takes logical terms to refer to formal features of the world, she holds that sentences get their logical form (including the associated modal properties)

from the underlying correspondence relations. But merely mimicking a *general* account of reference will not work for the expressivist, for it doesn't explain why only a *select* set of mental states inherit the modal properties of their contents. While the expressivist might seek to address this problem by adding some further feature, such a move is not promising. Consider, for instance, an expressivist who took this further feature to be some primitive aspect of mental states like belief and intention. Granted, this would allow him to maintain that only certain mental states are modally transmitting. But it would do so at a high cost. For starters, the move comes off as a particularly *ad hoc* solution. More significantly, it leaves him in a less theoretically sound position than his descriptivist rival: In addition to needing to mimic the descriptivist's account of correspondence, the expressivist must *also* add some further (perhaps primitive) feature in order to secure the needed selectivity. In light of this, there is no reason to think that the tools of Sher and Schroeder can secure an expressivist-friendly response to the modal test.⁴¹

(3.4) The Verdict. The investigation so far demonstrates that neither of the two general strategies for addressing the modal test provide expressivists with the tools needed to develop a viable account of the distinctive necessity that we associate with logical consequences. This conclusion is significant in a number of respects. For one, it draws out an as of yet unrecognized deficiency in the expressivists' explanatory

⁴¹ Might the expressivist do better by seeking to mimic the proposal of Wittgenstein 1921 (see note 35)? No. Wittgenstein's proposal answers the modal test by taking logical form to be an unanalyzable feature of the world. As such, it leaves unexplained the very features that the modal test demands one explain. Were no substantive account possible, this move to primitivism might be acceptable. But in the face of Sher's proposal, it is decidedly inferior.

resources—whereas the traditional objections focused on, among other things, *whether* expressivist accounts could mimic the truth-preserving nature of logical consequences, the argument here demonstrates that expressivists cannot explain a further feature of logical consequences—namely, why they’re *guaranteed* to preserve the truth. Moreover, we’ve learned that expressivists cannot just accept this failing as a further, minor limitation of their (self-admittedly) revisionist account. For starters, we’ve seen (§1) that this guarantee is the most significant feature of logical (as opposed to extra-logical) consequences. Moreover, the discussion of Sher’s proposal demonstrates that descriptivists are not similarly constrained. So not being able to capture it is devastating. The upshot is a new and forceful argument in favor of descriptivist meta-ethical accounts.⁴²

4. Might Expressivists Gain from a Skeptical Turn?

Given that expressivists can’t meet the modal test, a last hope would be to try and undermine its validity. Were expressivists able to do this, they could argue that their inability to address the modal test extracts little cost to the overall plausibility of their account. We can find the makings for just this sort of response in Blackburn 1987.⁴³ Following this line of argument, expressivists could note that the modal test

⁴² The argument here is significant in another respect. As we noted (§2), many of the existing critiques of expressivists’ alternative account of logic and inference just presume that a descriptivist account of logical form can be given. Seeing that this can be done by adopting Sher’s proposal provides these arguments with the further backing that they need.

⁴³ Some of the discussion in Gibbard 2003 suggests that he too might be inclined to a skeptical strategy. For instance, Gibbard (chap. 4) embraces Horwich’s conclusions that logic can be explained without

presumes that we have a good handle on the distinctive necessity associated with logical consequence. But they could then argue that this presumption is mistaken. This claim—if true—would allow the expressivist to set aside the modal test.

In Blackburn's hands, the argument for the second claim takes the form of a Euthyphro-style dilemma (53-4). Adapting it to the present context gives us the following: The modal test demands an explanation for why logical consequence has the distinctive necessity that it does. But providing such an explanation requires us to analyze the necessity of logical consequence in terms of some fact *M* about (say) language, rationality, or formal objects. Now either *M* is itself something necessary, or it isn't. If *M* is necessary, then we don't really seem to have explained the necessity of logical consequence. Rather, it appears we've just swapped one mysterious modal notion (the necessity of logical consequence) for another (the necessity of *M*). But if *M* is not itself necessary, then there's good reason to think that the necessity of logical consequence "has not been explained or identified, so much as undermined" (53). After all, if the necessity of logical consequence is grounded in something contingent, then it seems we should not conclude that the necessity of logical consequence has been explained, but rather that there's no such necessity! Expressivists could then conclude from this dilemma that the modal test relies on a mistaken conception of what the supposed necessity of logical consequence consists in—it incorrectly

appeal to a robust notion of truth. If he also endorses Horwich's defense of this conclusion, he is committed, with Horwich (1998: 71-6), to endorsing a skeptical strategy.

presumes that there is some *thing*, some feature of the world, on which the necessity of logical consequence rests.⁴⁴

There is something right in this move to skepticism about the modal test. For one, it would be reasonable to hold that the two proposals for addressing the modal test that appealed to features of thought and language fall to the second horn of Blackburn's dilemma. After all, these proposals seek to explain the necessity of logical consequence in terms of semantic notions like truth and satisfaction, or psychological notions of rationality and planning; and we've seen that these foundations are not stable enough to secure the modal profile that we associate with logic. However, this alone is not enough to save the expressivist. Sher's account, after all, does *not* explain the necessity of logical consequence in terms of something contingent. Rather, Sher answers the modal test by giving a substantive account of formal properties and objects that explains the *necessity* of logical consequence in terms of the *essential* features of things like sets and intersection relations. Does this mean her account falls to the first horn of Blackburn's dilemma? Does it mean that her proposal is deficient or trivial as Blackburn's argument would suggest? No. For one, we learn something interesting when we see that the modality associated with logical

⁴⁴ A couple of points: (1) Expressivists might find further support in the arguments of philosophers of logic (e.g., Etchemendy 1990, Brandom 2001) who are skeptical that the traditional strategies for meeting the modal test can be made to work (but see Sher 1996 and MacFarlane 2005 for reasons for thinking this skepticism is misplaced). (2) Expressivists that move to deflate the supposed of necessity of logic would need to provide an explanation of why we have mistakenly thought logic has a distinctive necessity. There are a variety of stories expressivists could tell here—for instance, Blackburn's (1987) projectivism or Brandom's (1994, 2001) pragmatism. So the success of the expressivists' skeptical strategy will turn, in part, on the plausibility of this explanation. Though I won't pursue it here, it is worth noting that the proposals of both Blackburn and Brandom are highly controversial (see, for instance, McDowell 1997; Rosen 1997, 1998; and Wright 1988).

consequence is of the same sort that we find in mathematics and set theory. But we also see that we can explain the necessity of logic in terms of the more familiar and tractable (structural) necessities of mathematics and set theory. This response shows the descriptivist how to blunt the first horn of Blackburn's dilemma. Furthermore, the explanatory power of Sher's account does much to undermine the hope for an alternative way to give the skeptical turn some bite.⁴⁵

5. Conclusion

If the arguments of this essay are sound, then we learn that descriptivists are much better positioned than expressivists to address the modal test. This is significant for several reasons. For starters, as we've seen (§3.4), it provides both a new argument against expressivist accounts of logic and inference, and a way of shoring existing critiques. More specifically, we see that descriptivists have—at least if they're willing to endorse certain claims about the nature and foundations of logic—the metaphysical and semantic resources they need to give a principled account of logical form, one that explains why logical consequences have the distinctive modal stability that they do. By contrast, the expressivists' commitment to grounding logic and inference in basic features of mental states leaves them unable to effectively mimic this descriptivists' solution. Given that this modal stability is the most distinctive feature of logical (as

⁴⁵ At this point, an expressivist might object that the account of formal properties is objectionably non-natural. However, there are a variety of strategies that a descriptivist can employ to make their account of formal properties consistent with naturalism. See, for instance, Linsky & Zalta 1995 and Maddy 1990.

opposed to extra-logical) consequences, this means descriptivist accounts provide a superior account of intuitive difference between the logical and the extra-logical.

While it's true that the argument of this essay requires the descriptivist to endorse controversial claims about the nature and foundation of logic, two points merit emphasis: First, though controversial, these claims are—as Sher's proposal draws out—plausible and explanatorily powerful. Second, as we have seen, the traditional criticisms of the expressivists' alternative accounts of logic presume an ability to distinguish between logical and extra-logical vocabulary. That is, they presume that there is an answer to the modal test. So the descriptivist must be prepared to defend a substantive account of the nature of logic anyway. Sher's proposal allows them to do that. The importance of the modal test runs deep—it tells us that vindicating the objectivity of ethics requires a descriptivist account of Factualism, not an expressivist one.

Chapter 2 is a reprint, with slight modifications, of my “Logic from Morals, Morals from Logic,” *Philosophical Studies*, forthcoming. Permission to reproduce this material has been granted by the copyright owner. The dissertation author was the primary investigator and author of this paper.

Chapter 3: Realism and Moral Error

If the arguments of chapter 2 are on target, a viable response to the skeptic must take the form of descriptivism. Descriptivism, recall, is the view that moral judgments are objective in that they express contentful states that describe or represent a distinctive realm of moral facts in virtue of predicating moral properties and relations to acts or individuals. As we will see, what one means by a “distinctive realm of moral facts” is the core of what’s at issue between descriptivists. But we can give some examples to help fix intuitions: for instance, one might maintain that moral facts are distinct in the sense of being *sui generis* properties; alternatively, one might take a moral property to be a special type of natural property—special in the sense of, say, its import for human well-being.

How do we proceed? Given the framework introduced in chapter 1 regarding how to understand the objectivity of ethics, we can turn to investigate questions of ontological structure. More specifically, are the properties and facts about which moral judgments purport to refer and describe ones that are a function of our thoughts, evidence, and institutions (constructivism), or are they in some way independent of these things (realism)? One fruitful way of making progress with regard to questions of ontological structure is to see how different versions of descriptivism propose to make sense of the objective features of our epistemological and critical practices. To that extent, this chapter formulates a second test—one concerning the critical notion of moral error. The project is to investigate whether—and at what cost—existing realist and constructivist views can make sense of moral error. Given how central the idea of

moral error is to the critical practices underling our discourse, any viable Objectivist account must be able to make sense of it.

Historically, the ability to make sense of error has been seen as a point in favor of realist accounts of our moral discourse.¹ In this chapter, I argue that this presumption in favor of realism is mistaken. Chapter 4 then explores the prospects for a constructivist account of moral error. Though we will see that existing constructivist accounts fail, we will also see that they do not exhaust logical space. By the end of chapter 4 we will have the material to formulate a further set of constraints that give shape to this newly found logical space. The remaining chapters will then investigate what sort of constructivist view might fit there.

1. Explaining Error: A Second Test

As we saw in chapter 1.4, our moral discourse exhibits a robust set of critical practices: we allow that our moral judgments may be incorrect; we think there can be genuine disagreements on moral matters even if we're agreed on all the non-moral questions; we can have and can improve our moral knowledge. These features make it common ground among Objectivists that our moral discourse supports a robust notion of fallibility. These observations point to the following constraint:

Fallibility Condition: A viable Objectivist account must provide an account of moral properties and facts that explains how our actual moral judgments can be wrong.

¹ This point is made by, for instance, Brink (1989: chap. 2), Sturgeon (1986), Shafer-Landau (2003), McDowell (1985), and Gibbard (1990).

In fact, acknowledging the Fallibility Condition shows why Objectivists cannot endorse simplistic subjectivist accounts of morality. These accounts identify moral facts with certain of our actual judgments or responses—e.g., wrong acts *just are* those acts that one judges disapprovingly. But the result is an account that can't make sense of moral error: Moral judgments and their truth makers cannot come apart since, by definition, they're the same thing.

But merely requiring the provision of an account of moral properties and facts that can give some explanation of error leaves too much open. For clearly, an account that held that a moral judgment was correct just in case a little angel whispers the answer in one's ear must be excluded. A second, two-part constraint—call it Tolerable Revision—seeks to address this.² As the angel example draws out, any account of moral properties and facts that requires positing such bazaar claims, or relies on radical psychological or cognitive capacities, would be—at best—unmotivated and ontologically extravagant. This gives us the first part of the constraint:

Credence Condition: General considerations of ontological and epistemological conservatism demand that we both have some reason to believe that there are moral properties and facts of the sort posited by a particular Objectivist account, and that those properties and facts comport with our best understanding of the world.³

² I formulate the two principles that follow to draw out their relevance to constraining our meta-ethical investigation. But it should be clear how the formulations in the text are mere specification of more general, topic neutral principles of theoretical investigation. See, for instance, Adams 1973, 1987, Brink 1989, Jackson 1998, Scanlon 1998, Shafer-Landau 2003, Smith 1995, Railton 1986a.

³ As should be apparent, this second test would be something that an expressivist account would also need to pass. Since we already have good reasons to reject expressivism, I will not pursue the prospects for a viable expressivistic account of moral error. I doubt that such an account would be possible. C.f., Sturgeon 1995, D'Arms & Jacobson 2000.

But the credence condition alone is not enough. An account of moral properties and facts must not do too much harm to our commonsense conception of morality. Even if we allow, as seems reasonable, that any account of what moral properties and facts are will almost surely have revisionary implications for the nature or content of morality, the core of our moral discourse and practice be preserved. This prompts the second part of the constraint:

Content Condition: A viable Objectivist account of moral properties and facts cannot involve unmotivated or severe changes to the content of morality.

Clearly, the Content Conditions is vague, but it must be in order not to beg questions. Even so, we can say more about both what counts as an “unmotivated” or “severe” change. A revision to the content of morality is *motivated* to the extent that the revision coheres with both (i) the account of moral properties and facts as a whole (including cohering with the marks of objectivity set out in chapter 1.4), and (ii) our best understanding of the world. A revision is *severe* to that extent that the resulting content is not recognizable as a moral system; a severe revision “changes the subject” in the sense that it’s not giving an account of morality, but of something else. Since the content condition remains vague, I will not place too much emphasis on it in the arguments that follow. Rather, its principle role will be to help us draw out more obvious failings.⁴

⁴ That said, some views (and their underlying presuppositions) place too much strain on our commonsense conception of morality and our understanding of the world (e.g., Nietzsche, Ayn Rand). These views are better understood as presenting skeptical, not Objectivist (in the non-Randian sense!), accounts of moral properties and facts. Others, Singer (1972) for instance, are harder to assess:

This pair of conditions provides a way to check if a particular account of moral properties and facts offered to explain moral error is plausible. We can now formulate our second test:

The Second Test: Any Objectivist account must (i) provide an account of moral properties and facts that shows how a moral judgment can be wrong [Fallibility], and (ii) do so in a way that is consistent with the content and credence conditions [Tolerable Revision].

This second test provides us with a further constraint on what counts as a viable account of the objectivity of ethics—one that will help us understand the ontological dimension of moral objectivity.⁵

2. The Independence of Moral Properties and Facts

The basic realist account of moral error states, roughly, that a given moral claim *m* is false just in case it fails to correspond with some mind-independent moral fact; *m* asserts that an act or individual has some mind-independent moral property that, in fact, it does not. Given the discussion of chapter 1.1, this means that realists have a two-fold obligation: First, there is the ontological project of providing an account of the ontological structure of moral properties that explains the manner in which they are independent of things like our thoughts, institutions, and evidence. Second, there's an existential project that aims to provide reasons that are sufficient

Arguably, Singer's revisions are well motivated (so it meets condition (i)), but I believe it is less clear whether they're too severe (so it's less clear whether it meets condition (ii)).

⁵ There is no guarantee that an account that passes the second test could explain other features of our critical practices that tell for the objectivity of ethics (e.g., the possibility of genuine moral disagreement, the possibility of moral knowledge). However, given the centrality of the notion of moral error to our critical practices, there is reason to think that an account that passes the second test will have the resources to make sense of these other features of our critical practices.

for thinking that the proffered moral properties actually exist.⁶ Moreover, notice that while these are distinct projects, they are not unrelated. As we will see, the realist's answer to the first requirement about ontological structure has significant implications for what they can say to convince us that there are moral properties of the kind they propose.

We can distinguish two general ways that realists have sought to explain what the ontological structure of moral properties is: identity strategies, and functionalist strategies. Identity strategies seek to specify what's distinctive of wrong acts (say) by specifying a set of mind-independent features that are necessary and sufficient for an act to be wrong. The result is an account of *what wrongness is*. Functionalist accounts, by contrast, pursue a more indirect method. They seek to provide an account of what wrongness is by looking to *what 'wrong' does*—that is, by specifying the role or function that 'wrong' plays within our moral discourse and practice. The expectation is that, given enough information about what the function of 'wrong' is, we'll be able to say both what the property of wrongness is, and why it's something mind-independent.⁷

In what follows, I focus on the dominant ways that realists have pursued these two strategies to answer questions about the nature of moral facts. This reveals that

⁶ Notice that these two requirements correspond to the second and third dimensions of objectivity (ontological structure and Realism, respectively) identified in chapter 1.1.

⁷ As will become more apparent below, the identity and functionalist strategies also typically differ across a second, epistemological dimension: though there are exceptions, identity strategies are generally pursued exclusively through *a priori* means, while functional accounts tend to draw on empirical data.

there is little reason to think that realists can provide an account that meets our second test.

3. Identity Strategies

Identity strategies seek to provide equivalence statements or definitions between the moral and the non-moral.⁸ Consider some examples:

Basic Utilitarianism: to be a right act =_{df} to be an act that maximizes aggregate well-being.

Simple Divine Command Theory: to be a right act =_{df} to be an act that is in conformity with God's demands.⁹

Given the identities, the outlines of an account of moral error is straightforward: one errs when one takes an act to be right (say) when, in fact, it fails to have the requisite feature—for instance, failing to maximize aggregate well-being. The challenge is to explain what it is about the right-hand sides of these definitions or identity claims that makes it reasonable to think there could be such properties. As we will see, there are a variety of ways that an advocate of an identity strategy might seek to do this, but I will argue that none proves successful.

⁸ Several comments are in order here. (1) The immediate focus of identity theories often concerns moral terms and concepts, and so they only derivatively address questions about the nature of moral properties. Furthermore, there has been much debate (often in conjunction with Moore's Open Question Argument) about whether the identity strategy should focus, in the first place, on concepts, or properties. For the purposes of the discussion that follows these issues will not be of concern. (2) There has also been extensive debate about whether moral identity statements *must* be fully reductive in the sense of identifying moral properties not just with non-*moral* properties, but with non-*normative* (e.g., natural) properties. Much of the debate turns on whether non-reductive accounts are suitably informative. Again, for the discussion that follows, it will not be necessary to delve into these matters.

⁹ The identity claims needn't be understood as definitions, but could understand instead as biconditionals. This distinction will not matter for the discussion that follows.

(3.1) Explicit A Priori Definitions. Some Realists seek to provide explicit definitions of moral terms like ‘wrong’ or ‘good’. These definitions purport to describe what is, as a matter of actual fact, common to all and only the wrong or the good things. Consider Fred Feldman’s statement of the project:

I mean to be searching for a suitably general statement of necessary and sufficient conditions for a life’s being good in itself for the one who lives it. ... Ideally, I would like to find a principle that would tell us when a life is better in itself for the one who lives it than some other life would have been. (2004: 13)¹⁰

A further distinctive feature of these definitions is that they are *a priori* definitions—we come to understand what the property of wrongness, or a good life, is by introspection, conceptual analysis, and thought experiments. What sanctions a claim of realism is that the content on the right-hand side of the definitions does not bottom out in claims about our thoughts, evidence, or institutions.¹¹

But, as we will see, advocates of explicit definitions find themselves seeking a precarious balance between the content and credence conditions: They need to define moral properties in a way that preserves the core content of morality; but in order for it to be plausible that there are properties of the proposed sort, the identities must be well motivated.

¹⁰ A few comments on Feldman’s project: (1) He is concerned to provide an account of what makes a life intrinsically valuable; he leaves open whether such a life would also be morally valuable. (2) He takes his project to be a project in normative ethics, not meta-ethics. However, given other work (e.g., Feldman 1978), it is not clear what alternatives meta-ethical positions other than explicit *a priori* definition Feldman could endorse.

¹¹ Much of the argument that follows applies with equal force to constructivist accounts (e.g., Lewis 1989) that seek to develop explicit definitions of moral and normative properties.

In order to draw out the magnitude of this challenge, let's consider Feldman's account.¹² Feldman (2002, 2004) starts with a version of basic hedonism, one that identifies value with net pleasure. But he quickly notes that such a definition does violence to our intuitive conception of the content of morality. For instance, basic hedonism cannot capture the plausible thought that promise keeping may be valuable even if it is unpleasant, or that the fruits of hard labor can be more valuable than equally pleasant events that result purely from good fortune. In light of these and other counterexamples, Feldman continually revises his definition of value until he arrives at the following preferred conception:

My view as a whole is roughly this:

1. The fundamental bearers of value are complex states of affairs of this form: S takes intrinsic attitudinal pleasure (pain) of intensity n and duration m in object P at time t , when S deserves to degree r to be taking that pleasure (pain) and P deserves to degree s to be the object of that pleasure (pain).
2. The desert adjusted value of such a state is a function of intensity, duration, truth, and pleasure (pain) worthiness of its object.
3. The double desert adjusted value of such a state is a function of intensity, duration, truth, pleasure (pain) worthiness of its object, and the pleasure (pain) worthiness of its subject.
4. The value (on the third scale) of a life is the sum of the desert adjusted values of the fundamental intrinsic attitudinal pleasure (pain) states in that life.
5. The value (on the third scale) of a world is the sum of the double desert adjusted values of the fundamental intrinsic attitudinal pleasure (pain) states in that world. (2002: 623-4)

¹² Adams (1973) presents a divine command theory version of the explicit definition strategy. However, he later rejects it and adopts a variety of moral functionalism. See Adams (1987) for the new account and an interesting discussion of why he's adopted it. We will consider the prospects of moral functionalism below.

One would hope that such a lengthy definition would not require significant revision to what we take to be essential to our notion of value. But it does. As Feldman himself admits, even this cumbersome account still requires him to accept certain highly implausible consequences—for instance, he notes that it will still be the case that relieving a sufficiently large number of persons from mild headaches is more valuable than saving an innocent from being tortured (2002: 624, 2004: 206). So Feldman’s preferred definition, at best, puts significant strain on the content condition.¹³

But even if we suppose that he could address this by adding more clauses to his definition, the resulting definition would violate the credence condition. Here’s why: For each new revision, Feldman must demonstrate that the addition of that revision increases the plausibility of (i) believing in the existence of such a (highly) conjunctive property rather than increasing the plausibility of (ii) denying that an explicit hedonistic definition of goodness is possible. But as the number of revisions needed increases, this bet grows increasingly implausible.

These problems provide strong reasons for doubting that an explicit hedonistic definition of goodness could meet the Tolerable Revision requirement. But—importantly—these concerns aren’t particular to hedonism; rather they reveal the familiar frustrations (some say futility) of seeking an *a priori* specification of an

¹³ The concern in the text demonstrates that Feldman’s preferred rendering fails to be *sufficient* to capture our intuitions about value. Similarly, the pursuit of fame likely meets Feldman’s definition, but—particularly when extreme—does not seem to make for a good life. But we can also question whether attitudinal pleasure is even a *necessary* element: A life of humility may well fail to bring one attitudinal pleasure, but why think that humble lives are less valuable than otherwise equal lives live without humility?

explicit set of necessary and sufficient conditions. Stephen Stich, among others, remarks that

No commonsense concept [including moral concepts like justice or rightness] that has been studied has turned out to be analyzable into a set of necessary and sufficient conditions. (1992: 250)

But we can say more about why this is so. The strategy of appealing to an *a priori* definition presumes that sufficient introspection or reflection on the nature of some moral property M will reveal the common set of features that unite all and only the legitimate instances of M-ness. This might be plausible to the extent that our thoughts about the property of being morally valuable (say) are either reasonably homogeneous or sufficiently amenable to revision, for under such conditions, reflection on seemingly disparate cases could, nonetheless, converge on a unique definition.

However there is no reason to think that either of these conditions hold for morality. For one, our folk intuitions about, for instance, what makes for a valuable life are an amalgamation of a rich and divergent set of religious, societal, and cultural themes about the nature and importance of things like modesty, respect, reciprocity, and the pursuit of individual pleasure.¹⁴ Furthermore, recent empirical work suggests that these divergent themes are, at least in part, wedded to deep features of our psychology—our feelings, emotions, and affective states. For instance, intuitions about modesty are driven by feelings of shame, intuitions about reciprocity are shaped by feelings of anger. This “psychological backing” makes the associated intuitions

¹⁴ Thomas Nagel (1979) makes this point nicely in his discussion of the various and incommensurable forms of value that are implicit in our moral discourse and practice.

difficult, if not impossible, to purge.¹⁵ In light of this, it's no wonder that a given *a priori* definition of a moral term proves vulnerable to counter-example: those terms are too much the product of disparate and psychologically rigid elements.

(3.2) Implicit A Priori Definitions. Considerations of this sort, have motivated some to seek an alternative version of the identity strategy—one that employs implicit definitions rather than explicit ones.¹⁶ Call this the implicit definition strategy (“IDS” for short). This alternative, rather than specifying an explicit set of conditions that is necessary and sufficient for being a moral property, moves in two steps. First, there is an *a priori* investigation that seeks to provide a “job description” that explains how moral concepts like ‘right’ function in our thought and talk. More specifically, the job description provides a characterization that implicitly defines our moral concepts in much the same way that the axioms of geometry implicitly define concepts like point, line, and triangle. The desired result is

a network of interconnected and interdefinable concepts that get their identity through their place in the network. We do not understand them one by one, but rather, holistically through their location in the network. (Jackson 1998: 130; c.f., 117-8, 139)

Second, the IDS advocate employs this network of concepts to identify whatever natural or non-natural properties (if any) fulfill this job description. Thus, the IDS

¹⁵ See, for instance, Haidt & Joseph 2004, Gill & Nichols 2008, D’Arms & Jacobson 2006.

¹⁶ Jackson (1998: 121ff, 150ff) might be read as being motivated by considerations of roughly this sort.

strategy takes moral properties be identical with the natural or non-natural properties that realize the associated implicit definitions.¹⁷

However, if these conceptual networks are to be capable of providing something that implicitly defines our moral concepts, they must have certain features. Most obviously, the resulting network must be (largely) consistent—otherwise we will not get coherent definitions. But the results must also be (sufficiently) specific. To draw out why an insufficiently specific network would be a problem, let's consider an example. As we have noted, the axioms of geometry implicitly define our geometrical concepts. But an interesting feature of the axioms of geometry is that they are *not* sufficiently specific in the sense that they do not circumscribe *unique* realizers for our geometric concepts. For instance, on a Euclidian realization of the axioms, the angles of a triangle will always sum to 180 degrees. But in non-Euclidian realizations, they can sum to more than 180 degrees. The upshot is that unless the advocate of IDS can demonstrate that his account meets these consistency and specificity requirements, we will have no reason to think there are (sufficiently) coherent and unique realizers for our moral concepts, and so no reason to think a claim to Realism is warranted.^{18, 19}

¹⁷ A couple of comments: (1) The indirect definition strategy is a type of realizer functionalism familiar from debates in the philosophy of mind and elsewhere. (2) The project of seeking a realizer for the implicit definition may be *a priori* or *a posteriori*. Naturalists like Jackson 1998 understand it as an empirical process; non-naturalists like Adams 1987, for obvious reasons, opt for an *a priori* characterization.

¹⁸ Witness Frank Jackson, an advocate of IDS, on this point: unless these requirements are met, the claim to Realism “will be more an act of faith than anything else” (1998: 29).

¹⁹ Though I will not pursue the point, it's worth noting that an advocate of IDS who seeks to vindicate a *realist* account of morality must also provide us with reasons for thinking that the realizers of his implicit definitions *must* include the right sort of mind-independent entities. Without this, the account would seem better understood as a variety of moral constructivism.

Since Frank Jackson provides, by far, the most worked out version of the implicit definition strategy, I will focus on his account. Jackson plausibly argues that the starting place for any *a priori* characterization of our moral concepts must be the whole of our folk morality—roughly, the complete collection of commonsense thoughts about things like right and wrong, good and bad (130-1). But while folk morality might provide us with a broad and familiar starting place, we've seen (§3.1) that it consists of a rich and divergent set of religious, social, and cultural themes. For instance, our folk morality has both retributivist and anti-retributivist elements: we think both that justice must be served when someone is wronged, and that at least some wrongs ought to be forgiven. The upshot—a point Jackson acknowledges (131-3)—is that there's good reason to doubt that folk morality will—on its own—be capable of providing us with sufficiently coherent implicit definitions.²⁰

To address the limitations with our actual folk morality, Jackson employs a strategy of reflective equilibrium on which we use process of *a priori* reasoning to weed out portions of our actual folk morality that fail to cohere in the right way with the rest (133). The trouble is that this approach threatens both the content and credence conditions. First, because the process of reflective equilibrium requires us to throw out parts of our actual folk morality, we must reject claims that, at least initially, we took to be quite plausible. The more we need to reject, the more revisionary the resulting account will be. Thus, with regard to the content condition, Jackson is making an

²⁰ While some (e.g., Thomson 1990, Kamm 1993) maintain that little if any revision to our folk morality is needed, such claims are hard to countenance in light of the arguments made above in §3.1.

unsubstantiated bet that the results of the reflective equilibrium process will not be excessively revisionary.

While we might be willing to take this bet, the difficulties Jackson's proposal has for the credence condition are too much to accept. The reflective equilibrium process requires a specification of what it is for a claim to cohere in the right way with the balance of our folk morality. Though Jackson does not provide an account of coherence, there are two options: he can understand coherence as a formal notion of consistency, or he can understand it in a more substantive way. But if coherence is taken to require just formal consistency, then the reflective process will not allow us to specify the refined network that defines our moral terms—there are just too many equally plausible, but incompatible ways of making an inconsistent set consistent. So Jackson needs something more substantive. But going this route requires him to specify what the relevant criteria for coherence are. Are we just concerned to preserve as much of our folk morality as possible? Do we instead want to ensure that certain “special” claims are to be retained? Something else? Not only does Jackson face the problem of there being a variety of intuitively plausible ways of giving substance to the notion of coherence, there's also the challenge of justifying whatever choice is made. Moreover, why think that such a justification is possible if all we can work with are the claims of folk morality itself?

These difficulties suggest that a defense of ISD like Jackson would do better by trying to supplement our actual folk morality with further constraints that would allow him to rule out certain claims and so arrive at a sufficiently refined set of

definitions. But what would these further constraints be? They can't be something with moral content, for on Jackson's account, everything with moral content has already been included in our characterization of folk morality. Nor can they be something formal—i.e., consistency—since we know that it alone isn't enough. So they must be things with substantive, non-moral content—e.g., scientific facts, claims about what the right worldview is (e.g., a theistic conception, a naturalistic one). The trouble is that to add substantive non-moral content of this sort is to give up on the implicit definition strategy. For one, moral concepts are no longer implicitly defined by the claims in the network of our folk morality, but rather by the combination of our folk morality and further *external* constraints.²¹ Moreover, if the account must rely on scientific facts—as it seems it must—the methodology is no longer (wholly) *a priori*.

In sum, we can now see that the IDS account relies on a dubious conception of implicit definitions—one that must violate either the content or credence conditions. But, importantly, the problem here is not unique to the particulars of Jackson's version of the implicit definition strategy. Rather, it's an endemic feature of any attempt to use implicit definitions to provide a non-stipulative, *a priori* account of the distinctive terms of a discourse.

(3.3) Unanalyzable, Sui Generis Primitives. The failure to secure viable explicit or implicit definitions of our moral terms has led some philosophers to

²¹ Here's another way to make the same point: IDS advocates like Jackson hold that their job descriptions provide an *analytic* account of moral concepts (Jackson (1998: chap. 6) is quite explicit about this). But the need to go beyond the network that is supposed to provide the analyticities shows that the desired definition cannot be provided.

conclude that moral properties must be identified with unanalyzable, *sui generis* moral properties.²² For instance, G. E. Moore claims that the property of goodness

is a simple, undefinable, unanalyzable object By what name we call this unique object is a matter of indifference, so long as we clearly recognize what it is and that it does differ from other objects. (1903: 21)

But because they maintain that these properties are *unanalyzable*, the arguments that seek to substantiate the claim to realism are invariably negative: The standard defense begins with an argument to the conclusion that no substantive account of moral properties is possible. The primitivist then uses this negative conclusion to argue that all existent substantive accounts of moral properties are fundamentally flawed. The defense concludes with a demonstration that primitive moral facts can explain the objective features of our moral discourse and practice.²³ Furthermore, it is in virtue of being understood as *sui generis* properties that these “primitivist” views count as varieties of realism.

But there is trouble. First, the primitivist’s argument relies on there *never being* a more plausible alternative account, and so it leaves her in a dialectically weak position—her account is plausible only to the extent that nothing better comes along.²⁴

²² In fact, one might see a move like this as providing an alternative way for an advocate of the IDS strategy to avoid the dilemma developed in §3.2.

²³ This line of argument can be found in Moore 1903 and Ross 1930. Yablo 1995 employs a more sophisticated form of the argument. Scanlon 1998 endorsed a version of this argument for non-moral value (more on this in chap. 4). Shafer-Landau 2003 and Wedgwood 2007 also take moral properties to be *sui generis*. However, they also maintain that moral properties are susceptible to some degree of analysis. As such, I take them to be better understood as varieties of functionalism and put them off for now.

²⁴ Cohen 2009 makes a similar point about primitivist accounts of color.

The primitivist might respond that hers is a defeasible commitment—one she'd be willing to give up were a viable alternative ever to appear. But this doesn't buy much comfort; she remains vulnerable to the development of a plausible alternative. Furthermore, I will argue in chapters 5-6 that there is, in fact, such an alternative.

More significantly, the primitivist cannot secure the credence condition. She is hard pressed to assure us that it is more plausible to countenance a realm of primitive moral properties that purport to explain the objective features of our moral discourse, than to follow the skeptic in maintaining that those "objective" features are misleading. For all that's been said—and all that it seems *could* be said—we have no reason to prefer the primitivist's "vindication" of the objectivity of ethics to the skeptic's revisionism. Nothing is gained if we avoid skepticism about the objectivity of ethics only by positing a realm of unanalyzable, *sui generis* properties—too much just goes unexplained.

To make this vivid, and to draw out a further way in which the primitivist strategy fails to meet the credence condition, let's explore what the primitivist can say about a novel moral issue like whether medical research on cloned human stem cells is morally permissible. Presumably, there is a moral fact of the matter here and so, given the credence condition, we can ask how might one go about discovering this fact (and thereby determine whether a given moral judgment is errant). Since the properties involved are *sui generis*, our access to them must be *a priori*. So, presumably, one gets traction on the fact of the matter here by some sort of reflection—one looks for a similar case to compare it to. However, there are a variety of cases that might count as

similar, yet they deliver incompatible results: If we deem the moral status of a cloned stem cell to be akin to that of a human, then one might be able to argue the research is impermissible.²⁵ But if it's more like a clump of animal tissue, then the research is permissible. The trouble is that since the property of being morally permissible is indefinable and unanalyzable there's *no principled way* of determining which of these two options tracks what's morally salient. So the primitivist lacks the resources to give an account of what the moral fact might be here that is capable of meeting the credence condition: if it is *not possible* to adjudicate between competing accounts of what the moral fact is, then we seem to have good reason to reject the primitivist's moral metaphysics. The upshot is that, barring some new alternative, there is no reason to think that identity strategies of any sort can substantiate a realist construal of moral properties.

4. Functionalist Strategies

(4.1) What Moral Functionalism Is. Functionalists hope to avoid the problems of primitivist accounts on the one hand, and explicit/implicit definition theories on the other, by seeking metaphysical middle ground. Against the definition theorists, they deny that moral properties can be given definitions that reduce moral properties to non-moral properties. But unlike the primitivist, they maintain that a substantive metaphysical account of moral properties is possible. The key, the functionalist

²⁵ This, of course, presumes that one has a settled position on the impermissibility of experimenting on humans. Furthermore, reflecting on debated about abortion, one might worry that the mere fact that the moral status of the stem cell is equivalent to that of a human is insufficient grounds for the conclusion that the research is impermissible.

maintains, is to ask not what moral properties are, but rather what role they play in our moral discourse and practice.

In virtue of this focus on the role of our moral concepts, functionalist accounts have much in common with the indirect definition strategy considered above (§3.2). In both cases, we get a handle on moral properties by specifying the role (or job description) of the associated concepts. However, unlike the indirect definition strategy, these roles aren't merely epistemic tools for finding the realizers that then get identified with the moral properties. Rather, the roles *are* the properties themselves. More specifically, moral properties are higher-order properties characterized by their distinctive functional roles. So, if act utilitarianism is true, the functional role of good would be something like 'being an act that maximizes net well-being'. This functional role would plausibly be realized by a variety of different lower-order properties (being an act of promise keeping, being an act of kitten rescuing...). But though these higher- and lower-order properties are *distinct*, a given higher-order property is metaphysically dependent on its lower-order properties in the following way: the higher-order moral property is *constituted by* (and so is not identical to) the lower-order non-moral properties in virtue of the fact that the higher-order property is *realized by* the lower-order ones. Furthermore, the functionalist allows that specifying the rightness role needn't be a wholly *a priori* process; one can appeal to empirical

considerations (e.g., historical case studies) to help determine (say) the plausibility of an act utilitarian construal of the rightness role.²⁶

By separating moral properties from the properties that realize them, and by endorsing an *a posteriori* account of how we specify things like the rightness role, the functionalist seems well positioned to avoid the difficulties that frustrated the implicit definition strategy. However, the moral functionalist still faces the challenge of demonstrating that she can meet the credence condition, for positing functionalism *alone* does not give us reason to accept a realist account of moral facts.²⁷ This is because we can give a functionalist account of paradigmatically constructivist properties like the property of being in accordance with club rules, or the property of being a birthday party. More specifically, the property of being in accordance with club rules is plausibly a high-order property constituted by lower-order properties like being an act of wearing one's club badge and being an act of deference to senior club members. But—importantly—since these realizers are a function of our social *institutions*, the property of being in accordance with club rules cannot be a realist property.

²⁶ Moral functionalists can also be distinguished in terms of whether or not they endorse a naturalistic worldview. The *naturalistic* functionalism of Boyd 1988, Brink 1989, Jackson 1998, and Sturgeon 1985b seeks to make a metaphysical connection between moral and natural properties. The *non-naturalistic* functionalism of Adams 1987, Shafer-Landau 2003, and Wedgwood 2007 denies that limiting oneself to a base of natural properties is enough. These views also diverge in their moral epistemology: naturalistic accounts typically endorse some variety of *a posteriori* coherentism, non-naturalists an *a priori* foundationalism.

²⁷ I do not mean to imply that non-reductionists think that the move to role functionalism alone is enough to establish realism about moral facts. Typically, they do not.

In response, functionalists reply that understanding of moral facts in a realist fashion is plausible because so doing *best explains* our moral discourse and practice:

[The] features of actual and possible moral inquiry are hard to understand on [constructivist or skeptical] assumptions and are much easier to understand on realist assumptions. Realism, and realism alone, provides a *natural* explanation or justification of the way in which we do and can conduct ourselves in moral thought and inquiry. (Brink 1989: 24, original emphasis)

In its most potent form, the argument employed by Brink and other moral functionalists has two parts:

- (i) Demonstrating that we have independent reasons for thinking that moral properties can't be given a plausible constructivist rendering.
- (ii) Providing positive reasons for thinking that moral properties merit a realist construal.²⁸

A comprehensive review of the arguments the functionalist might offer against constructivist views is not possible. So to get a gauge on the functionalists' ability to achieve part (i) of their best explanation argument, let's focus on two arguments that they have taken to be particularly problematic for constructivist theories.

(4.2) The Functionalists' Critical Argument. David Brink holds that the "chief objection to any form of constructivism [is] its identification of truth and justified belief" (1989: 31).²⁹ More specifically, he begins by noting that the *truth* of a moral belief *m* is independent of one's *justification* (actual or ideal) for believing *m*. He then argues that, since constructivists must wed themselves to an irrealist conception of

²⁸ See, for instance, Brink (1989: 12, 31-5, 182-97); Shafer-Landau (2003: chaps. 2, 4, 8).

²⁹ C.f., Sturgeon 1986.

truth,³⁰ they cannot make sense of the distinction between truth and justification.³¹ The significance of this failure gives us good reason to reject constructivist accounts. But this argument can't be right, for it generalizes in implausible ways. To see this, consider our discourse concerning social etiquette. If we're constructivists about social etiquette, we're—according to Brink—wedded to an irrealist account of truth. As such, we will be unable to distinguish the justification of the belief that forks and napkins are placed on the left from its truth. But if we follow the reasoning of Brink's argument against moral constructivism, this failure gives us good reason to reject a constructivist account of etiquette. But clearly something has gone wrong—for etiquette is a paradigmatic example of a constructivist discourse. Seeing that Brink's argument generalizes in implausible ways points to what the problem is. The argument gets its force, not from the claim that the moral constructivist can't distinguish justification from truth, but from the presumption that our moral discourse must be given a realist rendering. But that assumption just begs the question.

Russ Shafer-Landau (2003: 42-3) provides a different general argument against constructivism. He maintains that a Euthyphro style problem “represent[s] a serious *prima facie* difficulty” for any constructivist moral theory. The idea is that the starting

³⁰ A realist about truth seeks to explain the truth-making connection between a sentence and the fact that it's about in terms of some mind-independent relation—e.g., a substantive notion of correspondence. The irrealist denies that there is any such mind-independent relation; she seeks to explain truth in terms of, e.g., the (ideal) justification, or pragmatic value, of one's belief set.

³¹ One might balk here—for some constructivists explicitly claim to use realist notions of truth (e.g., Wong 1984), while other views seem consistent with it (e.g., Harman 1975, Scanlon 1998). So if Brink's argument is to undermine *all* varieties of constructivism, it must be supplemented to show the impossibility of combining constructivism and realism about truth.

point for the construction is either moralized, or it's not. If it is moralized, then we're not dealing with a constructivist account of morality—all the moral elements are just built in from the start. If it's not moralized, the trouble is that “there is no reason to expect that the principles that emerge from [the] construction process will capture our deepest ethical convictions, or respect the various platitudes that fix our understanding of ethical concepts” (42).³² However, second horn of Shafer-Landau's dilemma can be resisted. For starters, while it's reasonable to demand that any meta-ethical view preserve the core content of morality (the content condition above asks as much), we must also allow that at least some revision will be necessary. More importantly, we must assess a meta-ethical account's ability to capture the content of morality in light of the plausibility of the metaphysical and epistemological assumptions it employs—surely meeting the content condition while violating the credence condition doesn't count for much.³³ Finally, as we will see in chapters 5-6, constructivist accounts that take our actual moral beliefs and principles as their starting place are well positioned to capture the core content of morality and so blunt the second horn of Shafer-Landau's dilemma.³⁴

In light of the serious problems with these two anti-constructivist arguments we have reason to doubt whether the functionalist could provide a strong, *general*

³² Similar argument are given by Brink (1989, 2006) and Gauthier (1985: 4-6).

³³ For this very reason we should be wary of the meta-ethical non-naturalism that Shafer-Landau advocates. More on this below.

³⁴ Scanlon (1998) seems inclined to deny the first horn of Shafer-Landau's dilemma. While I think Scanlon's argument fails, I put off discussion of it until Chap. 4.

argument against moral constructivism.³⁵ Realizing this draws out the importance of the second phase of the functionalist's strategy (part (ii) above): The success of their best explanation argument rides on their ability to demonstrate that we have substantial, positive reason for endorsing a realist account of moral properties. It's worth emphasizing that since the functionalist makes the claim that realism about moral properties best explains our discourse, the burden is on her to provide a convincing case that this is so.

(4.3) The Functionalists' Positive Case: The Analogy Argument.³⁶ The most plausible positive argument to the conclusion that we have reason to endorse realism about moral properties takes the form of an argument from analogy. Brink sums the strategy nicely:

the metaphysical and epistemological commitments of moral realism are very similar to, *and so no less plausible than*, those of realism about commonsense physical theory and the natural and social sciences. (1989: 12, emphasis added)³⁷

Pushing beyond Brink's gloss, we can specify the central inference of the best explanation argument as proceeding like this:

- (1) There are certain features of the natural and social sciences that, on the whole, give credence to a realist construal of the underlying properties of those discourses (call these the "telling features").

³⁵ Importantly, this is not to say that there are no strong arguments against *particular* versions of constructivism. As will become apparent in chapter 4, I believe there are many such arguments.

³⁶ The argument that follows is a shortened version of one developed in Kurth ms. 2.

³⁷ Similar commitments can be found in, for instance, Boyd 1988: 181-2, 204-5, Railton 1986a, and Sturgeon 1985b, 1986.

- (2) Moral discourse, on the whole, displays similar telling features.
- (3) Therefore, we have good reason to endorse realism about the properties underlying our moral discourse.

Premise (1) is a claim about the features of scientific methodology that, when present, sanction a realist construal of the underlying properties. These features—the telling features of science—include things such as: (i) The sciences support notions of error and improvement. (ii) They allow for the possibility of substantive disagreement. (iii) Inquiry within science is viewed as reasoned (*a posteriori*) deliberation. (iv) Scientific properties and facts contribute to explanations of certain states of affairs. (v) Claims about the underlying properties of a science admit of predictive testing. And (vi) scientific properties and facts provide feedback—they have shaping influence on our thoughts about the area of science in question.³⁸

We can get a picture of how realists deploy the analogy argument by looking at Peter Railton’s version of it.³⁹ Railton begins by noting that the presence of rich explanatory power tells for a realist construal of a discourse. For instance, he notes that

³⁸ These telling features can be found, in various forms, in Brink 1984, 1989, Sturgeon 1985, 1986, 1994, Railton 1986, 1989, Boyd 1988, and Bloomfield 2001. It is worth mentioning that the “telling features” cited in the text are those that naturalistic moral realists appeal to in order to vindicate their *realism*. These features can be distinguished from other features of our moral discourse—e.g., declarative syntax, assertoric form, embedding within unasserted contexts, etc.—that realists point to in order to vindicate a more general claim to moral *cognitivism*. Since my concern is with the debate between realists and constructivists, in what follows, I will focus on the features that tell for realism.

³⁹ See Railton (1986a, 1989). Though there would be variations in the details, a similar account could be told for the other versions of naturalistic moral realism like those referenced in note 38. Brink’s account may be somewhat of an exception—while he does appeal to the telling features to make a case for realism over constructivism, he uses these features more often as part of an argument against non-cognitivism.

[w]e are confident that the notion of chemical valence is explanatory because proffered explanations in terms of chemical valence insert explananda into a distinctive and well-articulated nomic nexus, in an obvious way increasing our understanding of them. (1986a: 17)

The driving thought here is that it's *because* of this rich explanatory power that we're inclined to be realists about the property of valence. In fact, the case for realism about valence grows stronger once we notice that it exhibits other telling features. For instance, our scientific account of what valence is entails that water molecules form stronger ionic bonds than do molecules of ethanol. This allows us to predict that a volume of water will have stronger surface tension than ethanol, and we can run experiments to confirm this prediction. We can also see that chemical properties provide feedback in the sense that these properties have a shaping influence on our perceptions, thoughts, and action. Our thoughts about valence and bonding are shaped by our perceptions: because we *see* that the water-spider can walk on the lake surface, but sinks on the oil slick, we *come to think* that different fluids have different surface tensions.

Having established that the presence of rich explanatory power via the telling features warrants realism, Railton turns to argue that moral properties and facts work in the same way as the properties and facts of chemistry. To do this, Railton begins by developing a functional gloss of the moral property of rightness—right acts are those that would be approved of from the social point of view. But, importantly, Railton's functional gloss is a stand in for a more specific, functionally characterized, causal mechanism—the wants/interests mechanism. The wants/interests mechanism is

(roughly) a set of causal ties that both hold among groups of individuals, and that functions to bring the group's (subjective) wants in line with their (objective) interests (22-3).⁴⁰ With this functional characterization in hand, Railton then argues that his functional proposal has the same explanatory power that we find for scientific properties like valence.

To demonstrate this explanatory power, Railton appeals to the telling features. For instance, he notes that his account entails that when the interests of a group are compromised, the wants/interests mechanism will incline the society toward unrest. This means that Railton's account predicts that social-political arrangements that would not be approved of from the social point of view will tend to spawn discontent. He then shows that the histories of oppressed societies provide evidence that confirms this prediction. Railton also demonstrates that rightness and wrongness as he characterizes them participate in informative explanations. For instance, to explain the storming of the Bastille, we cite the fact that the political arrangement in France was wrong. He also maintains that his account of rightness gives a prominent role to feedback mechanisms: The discontent generated by, say, an oppressive political arrangement can produce feedback (through the wants/interests mechanism) that "promotes the development of norms that better approximate" what would be approved of from the social point of view (24). Railton again points to historical cases—e.g., the suffrage and civil rights movements—as evidence for these social-

⁴⁰ Boyd 1988: 203-4 and Brink 1984: 121-2 make similar proposals.

psychological feedback mechanisms leading to better moral/political arrangements. Railton takes the upshot of his investigation to be that we have good (*a posteriori*) evidence for the claim that moral properties exhibit the same telling features as the sciences. So, in virtue of the analogy argument, he concludes that we're warranted in endorsing realism here as well.

(4.4) The Functionalists' Dilemma. A common objection to arguments like Railton's holds that morality is importantly disanalogous with science. In its most sophisticated form, the objection first notes the high degree of abstraction at which the realists' analogy operates. It then maintains that the resulting examples of morality displaying the telling features are just too thin, too trivial, too platitudinous to do the needed work. But these realists seem to have a ready reply. Railton, for instance, considers the complaint that the examples that he provides in defense of his moral realism are insufficiently robust. His response is two-fold: First, he acknowledges that his account has but a "breezy plausibility"—by which he seems to mean something like 'an intuitive fit within our larger picture of the world' (15). But, second, he denies that this is a problem because we're willing to endorse comparably "breezy" accounts in the sciences. For instance, he maintains that his evidence for the claim that moral properties display feedback mechanisms "has about the same status...as the more narrowly biological argument that we should expect the human eye to be capable of detecting objects the size and shape of our predators and prey" (15). Given this

similarity, Railton maintains that there's no reason to think the analogy is too superficial to support a claim to realism.⁴¹

However, while replies like this indicate that the realists' can provide examples where morality displays the telling features in ways that are comparable to what we find in the sciences, it also reveals a dilemma. The question that gets the dilemma going asks how robust the analogy between science and morality must be in order for it to be capable of doing the work the realist needs it to do. In the sections that follow, I will demonstrate that if the standard for robustness is something like Railton's "breezy plausibility," then paradigmatically constructed discourses like fashion display the telling features in manner that is on par with what we find for our moral discourse. So the analogy with science fails to support moral realism. I then consider whether the realist might be able to provide a more robust standard. Here we will see that the most promising alternatives are unable to deliver the needed analogy with science. In fact, learning why these proposals fail demonstrates that other proposals are also likely to fail. So again, the analogy argument fails to vindicate moral realism.

(4.5) The Challenge from Fashion. Fashion offers a sharp contrast with science. While science is the paradigm of a descriptive and realist discourse, fashion—as a discourse that is both prescriptive and constructed—marks out the other extreme. That fashion is both prescriptive and constructed is readily apparent given our discourse and practice: when we say that a style is fashionable, we (typically) intend to

⁴¹ See Brink 1989: chap. 7, Sturgeon 1985, and Bloomfield 2001: chap. 1 for the makings of similar lines of response.

commend it; we take fashion to be a paradigmatic example of a human construction because facts about what's fashionable are so intimately tied to our thoughts, tastes, and conventions.

I believe this picture of our fashion discourse captures our primary use of predicates like 'is fashionable' and 'is chic.'⁴² However, though fashion properties and facts are constructed, they can still support the telling features with the same breezy plausibility that we find for moral properties and facts. Consider the explanatory/predictive power and feedback mechanisms on which Railton and others rely: We can explain why George Clooney made *People* magazine's annual "best dressed" list by citing the fact that he is a fashionable dresser. And, given that he dresses so well, we can predict that he will make the list again next year.⁴³

Furthermore, the accolades and condemnations that are elicited by instances of chic

⁴² As noted in the text, our primary use of terms like 'fashionable' and 'chic' aims to pick out constructed properties and facts. However, there are a couple of secondary uses where one might think that a realist construal is plausible. Moreover, because my argument builds on a notion of fashion that is constructed, it will be worthwhile to briefly explain why these secondary uses are not appropriate for the discussion that follows. First, there may be a secondary use of 'fashionable' that is more or less equivalent to judgments of beauty (e.g., when we talk of a style that is "timelessly fashionable"). While this secondary use might have some claim to a realist construal, that claim would be at least as controversial as the associated claim to realism about the beautiful—and likely more so. More importantly, this fashion-as-beauty use is ill-suited for capturing central features of our fashion discourse: as noted in the text, predicates like 'is fashionable' and 'is chic' are typically used to pick out properties that are intimately tied to *fluid* social conventions and individual *tastes*. As such, a realist construal of them is generally thought implausible (see Railton 1997 for a similar point). Second, there is a wholly sociological/descriptive use of 'is fashionable' that tracks things like what individuals or cultures regard as fashionable, and how fashion norms evolve and get transmitted. But because this notion is descriptive, it fails to capture the normative function that is central to the core of our fashion discourse. While there are interesting questions about how this secondary descriptive use is related to the primary prescriptive one, I will not take that up here.

⁴³ It's worth noting that fashion properties and facts pass the counterfactual test that some naturalistic realists (e.g., Sturgeon 1986, Railton 1986a, Brink 1989) take as evidence for the robustness of a particular explanation or prediction. For instance, the claim 'were Clooney not such a fashionable dresser, he would not have made all the stylish clothing choices that he did' seems just as true as Sturgeon's claim 'were Hitler not morally depraved, he would not have done all the evil that he did'.

and unfashionable dress provide feedback that can shape thought and action. For instance, given that fashion functions to express individuality and group membership, comments and criticism from friends can influence the style of dress that one deems fashionable.⁴⁴

But our fashion discourse doesn't just support substantive explanations, predictive testing, and feedback mechanisms. It also supports notions of error and improvement. Notice, for instance, that we say things like "Bob used to have terrible style [error], but since he started dating Jessie, he's really learned how to dress more fashionably [improvement]." Such talk is not merely metaphorical. First, fashion allows for genuine fallibility since one can fail to accurately perceive the fashion conventions of a given group. Moreover, because we can get better at identifying both fashion conventions, the above model can explain how our fashion judgments can improve.

In addition to error and improvement, fashion also supports substantive inquiry, deliberation, and disagreement. Because it can be difficult to figure out what the relevant fashion conventions are, and because one might not be sure what sort of statement a particular outfit might make, it's not surprising that we find inquiry and deliberation in our fashion discourse. Not only do we ask for advice about what to wear, but we also tend to give reasons to substantiate our critical judgments or to justify the styles we've adopted (e.g., Coco wears it, so it must be chic). Moreover, to

⁴⁴ It's significant that this shaping needn't be conscious: I might deliberately decide to change the way that I dress because I believe your critical comments are correct; but your comments might also cause me to unknowingly start to mimic your style. C.f., Brink 1989: 188-9.

the extent that fashion facts are facts about how to best balance group membership and individual expression, it also makes sense that we find substantive disagreement about what's chic.⁴⁵

Clearly such results are highly problematic for the moral realist's analogy argument: If the properties and facts of paradigmatically constructivist discourses like fashion display the telling features just as well as those of morality, then the analogy with science does nothing to help establish the truth of moral realism. That fashion turns out to be on all fours with morality shouldn't be surprising. Given the high level of abstraction at which we're working, breezy plausibility is easy to get. It can be secured by any discourse—morality to fashion—that plays a broad, regulative role with regard to human social interaction.

(4.6) The Search for a More Substantive Standard. If the analogy argument is going to work, the realist needs to find a standard that is more substantial than breezy plausibility. However, there's good reason to think this search must fail. In order to establish a more robust standard, the realist must (i) identify features of explanation, prediction, etc. the having of which would help substantiate a claim to realism, and (ii)

⁴⁵ Consider the following exchange from the *New York Times* blog coverage of the Fall 2009 Prada fashion show (Horyn 2009):

Post 1: “[I]f Miuccia Prada is doing fishing waders in her show, you know she's not just doing any old waders. And this is not just any collection. I loved the equal treatment, the equal value, she gave to the [fishing waders] and to posh fur, or to glossy velvet and practical country tweeds. ... What made this collection work so well—energetically, brilliantly—is that Prada applied the thinking across the board. Everything was consistent”

Post 2: “I see no designs at all...just mad-cap random styling of pre-existing items.”

Notice that not only is there disagreement about whether Prada's fall collection is fashionable, but both commentators offer reasons to substantiate their claims (“equal treatment” of different elements in a “consistent” manner vs. “mad-cap random styling of pre-existing items”).

demonstrate that these features are had by morals and science, but not fashion. In what follows, I will consider two candidates that figure prominently in the literature: the potential for belief independent explanations, and the possibility of robust forms of error and disagreement. As we will see, both proposals fall short. Moreover, seeing why they do will reveal that alternative proposals are likely to fail as well.

(4.6.1) Belief Independent Explanation. The first possibility appeals to belief independent explanations. In explanations that tell for realism, the facts in question have explanatory power even when nobody has any beliefs about those facts: Facts about chemical valence explain why we observe the water-spider sinking once it gets to the oil slick. But they would do so *even if* we had no antecedent beliefs about ionic bonds, surface tension, and the like. This capacity for belief independent explanation is part of what inclines us to realism about valence. So the realist might seek to revive the analogy argument by showing that morality exhibits a comparable belief independence.

Let's grant for the moment that fashion explanations are not belief independent in the relevant sense. To demonstrate that moral explanations are belief independent, realists draw on thought experiments. The stock example focuses on explaining why an unjust society will have a tendency for unrest. The realist maintains not only that moral facts about the unjustness of the society explain why it has the potential for unrest, but that we can explain this tendency for unrest even on the assumption that no one there believes that their society is unjust. Consider Railton's version:

Suppose that a given society is believed by all constituents to be just. This belief may help to stabilize it, but if in fact the interests of certain

groups are being discounted, there will be a potential for unrest that may manifest itself in various ways—in alienation, loss of morale, decline in the effectiveness of authority, and so on—well before any changes in belief about the society’s justness occur, and that will help explain why members of certain groups come to believe it to be unjust, if in fact they do. (1986a, 23; c.f., Sturgeon 1986, Brink 1989: 189)

But in light of thought experiments like Railton’s, we should ask whether we really have an example of a belief independent explanation.

With regard to the first question, notice that Railton presumes he can demonstrate that his explanation of the generation of unrest in the society is belief independent by telling a story in which no one in that society has the particular moral belief that the society is unjust. But why think that eliminating just this *particular* moral belief—namely, that the society is unjust—is sufficient to establish that morality supports belief independent explanations? To see why more needs to be said, notice that were the individuals in the society to have certain *general* moral beliefs—e.g., general beliefs that their society could be morally better, or un(der)specified beliefs that there’s something morally amiss in their society—we would also be able to explain the tendency toward unrest.

Recognizing this reveals that establishing belief independence requires more. In addition to claiming that the individuals lack the particular moral belief that their society is unjust, realists like Railton must also maintain that the individuals in the unjust society have no associated general moral beliefs (or, if they do, they’re not implicated in generating the unrest). More specifically, the argument requires that it be possible for moral unrest to develop in a society *even if* the individuals there have

extremely impoverished belief sets—they have *no* specific beliefs about the unjustness of the society, *no* general beliefs about the comparative value of the social arrangement of their society, and *no* un(der)specified beliefs that something is wrong in their society.

But why think that unrest could develop among individuals with such impoverished belief sets? Granted, there's nothing incoherent in the thought of a wholly ignorant but discontent society. But whatever plausibility this conceptual point might have, it is of little help to the realist. After all, realists like Railton are trying to establish that something like a wants/interests mechanism actually exists. So an example that merely establishes a *conceptual* possibility won't do. We need to see that it is *psychologically* possible for beings like us to have a tendency for unrest in the absence of any specific beliefs about the unjustness of the society, any general beliefs about the comparative value of the social arrangement of their society, and any un(der)specified beliefs that something is wrong. But such psychological claims are hard to countenance. Consider, for instance, the alienation that Railton refers to in the quoted passage. Presumably, the thought here is that in unjust societies, members of the oppressed groups become alienated in the sense that they come to feel that they have lost control of their lives. But notice that for one to become alienated in this way seems to require one have at least some sense of *entitlement*—e.g., a general belief that one has a right to control certain aspects of one's life. So it's hard to see how a belief independent explanation could be possible—having a tendency for unrest

appears to require having general moral beliefs about the social arrangements of the society.⁴⁶

(4.6.2) Deep Error. A second option that the realist might appeal to focuses on the nature of scientific error and disagreement. More specifically, the hope would be to vindicate the analogy argument by identifying forms of error that science and morality display but that fashion does not. This move has promise: science and morality seem to involve special forms of error—forms that would seem to support a claim to realism.⁴⁷ Consider: while constructivist discourses like fashion make some distinction between the *truth* of a belief and the *justification* that one has for it, they must hold that there is at least some tie between the two. After all, the conventionalist takes the facts in question to be constituted by some (perhaps idealized) set of beliefs, conventions, or responses. This is interesting because it seems to limit the forms of error that the constructivist can accommodate. In particular, a conventionalist about (say) fashion can hold that one's fashion judgment is incorrect just in case it fails to correspond with the set of beliefs, conventions, or responses that are constitutive of the relevant fashion facts. But while the conventionalist *can* explain fashion errors that result from not having one's fashion judgments conform to the relevant set of beliefs/conventions/responses, he *cannot* explain the possibility that one could be

⁴⁶ Railton's own discussion seems to concede as much. He notes that a "common theme" in the empirical work on social unrest is that societies tend to become discontent, not because they see their society as unjust, but because they see their "customary entitlements" as being threatened (40n33). I take the pervasiveness of this pattern to support the claim in the text that social unrest requires individuals to have (at minimum) general moral beliefs about what they are entitled to.

⁴⁷ Arguments of this sort are made by, for instance, Sturgeon (1985), Railton (1986: 4-5), Brink (1989: 31, 88-9), and Bloomfield (2001: 3-23).

mistaken about the beliefs/conventions/responses that make up that set. This sort of deep error—error which allows that *any* of one's views might be mistaken—only seems possible for realist discourses like science and morality that take the truth of a belief to be *completely independent* of the justification that one has for it.

The realist is right to notice that our intuitions about error reveal important differences between science and morality on the one hand, and fashion on the other. But he's wrong to conclude that these differences are to be explained by taking science and morality to be realist discourses and fashion a constructed one. Once we investigate what it is that makes deep scientific error seem possible, we see that there is little if any support for the crucial premise of the realist's argument.

As we noted, to say that a discourse allows for deep error is to say that the truth of one's beliefs is completely independent of the justification that one has for them. So what is it about science that makes us think that it possesses this robust independence? Well, for starters, science is full of examples of claims and theories that were thought to be correct at one point in time, but were later completely abandoned (e.g., Aristotelian medicine, phlogiston theory, the Ptolemaic model of the universe). Because we have these examples, we have some indirect support to the thought that the truth of our scientific beliefs could be completely independent of the justification that we have for them. But claims about the independence of science gain more direct—and so more compelling—support from the observation that science supports belief independent explanations. After all, the ability to explain scientific

phenomena without *any* appeal to our scientific beliefs is indicative of the *complete* separation between truth and justification that makes deep error possible.

Given the nature of the realist's analogy argument, we should expect these two phenomena to be nicely replicated in our moral discourse, but not our fashion discourse. But this is not what we find. For one, we have seen that neither morality nor fashion supports belief independent explanations. This is significant. The fact that we must appeal to moral/fashion beliefs in order to explain moral/fashion happenings indicates that there is an essential connection between truth and justification in these discourses. So they lack the independence necessary for deep error to be possible. While I take this point to tell strongly against the possibility of deep moral error, it's worth noting that the realist gains little from an appeal to the thought that morality, like science, has examples of claims and theories that have been completely abandoned. This is because ideal observer varieties of moral constructivism can also explain out intuitions about how moral inquiry has led us to abandon various claims and theories. So the argument offers no support for realism. The upshot is that the appeal to error and disagreement—like the appeal to belief independent explanation—fails to provide the realist with a way to revive the analogy argument.

While there may be additional strategies realists could employ to revive their argument, we have good reason to think they too must fail. For one, we have seen that the three most prominent proposals in the literature have fallen short. Moreover, in examining them, we have identified a series of important differences between morality and science—only science seems able to support belief independent explanation and

deep error. These differences indicate that any new strategy the realist might propose will be incapable of securing a suitably robust analogy between morality and science.⁴⁸

(4.7) The Verdict. We have seen that if the moral functionalist's analogy argument is to support his moral realism, he must provide us with an account of the manner in which our moral discourse is supposed to be analogous with science. But we've learned that answering this question puts the realist between the horns of a dilemma. If establishing the analogy with science only requires a thin degree of similarity—a breezy plausibility—then, though the analogy holds, it fails to support a claim to realism. But attempts to identify a more robust standard fall short—either they also fail to warrant a claim to realism, or they're something that our moral discourse doesn't support. So we should conclude that, despite its prominence, the analogy argument fails to vindicate naturalistic moral realism. But when we combine this failure with the functionalists' troubles in providing a general critique of

⁴⁸ A final realist strategy: The realist could respond, not by tackling the second horn of the dilemma, but rather by denying that there is a dilemma in the first place. In particular, were one able to make a plausible case for fashion realism, then the tension in the analogy argument that the dilemma seeks to exploit would disappear. In fact, it seems that a plausible case for fashion realism can be developed on the model of aesthetic value proposed in Railton 1997.

Why this strategy fails: As discussed above, a realist account of our primary use of predicates like 'is fashionable' and 'is chic' fits poorly with the robust connections that our judgments about what's fashionable have with fluid social conventions and individual tastes. So there's little reason to think that this strategy would work. In fact, this very point is nicely illustrated by Railton's proposal: He acknowledges that his account of aesthetic value is importantly different from his account of moral value in that only for the former are the values in question sensitive to culturally variable factors (1997: 116, 124). In light of this, it is difficult to see how his account of aesthetic value could be used to make a case for fashion realism. Moreover, it's also worth noting that Railton does not present his account of aesthetic value as an account of *realism*; rather, he presents it as an account that is capable of securing robust forms of objectivity. In so doing, I believe we might understand him as acknowledging a central theme of this essay—namely, that securing robust forms of objectivity does *not* require a realist metaphysics.

constructivism, and in establishing that moral discourse is really like paradigmatically realist discourses, we have a powerful argument to the conclusion that the functionalist cannot substantiate his realist understanding of moral properties. Moreover, to the extent that these naturalistic realists are right that their account provides best hope of addressing the traditional metaphysical and epistemological worries that undermine alternative realist proposals,⁴⁹ then the dilemma carries force not just against view like those of Railton, Boyd, and Brink, but moral realism more generally.

5. Conclusions and Implications

Realists hope to account for error by taking moral properties to be independent of our thoughts, evidence, and institutions. But we've seen that this strategy meets the Fallibility Condition only if it violates the credence condition. However, reviewing how the various realist accounts fail reveals some interesting conclusions. Specifically, what we see is that no Realist account is able to meet the credence condition either because the account proffers a bazaar or incoherent account of what moral properties are, or because the account provides no reason to favor a Realist construal of moral properties over a constructivist rendering. This is doubly significant. For one, it means that we have no reason to believe that the objectivity of ethics is grounded in some realm of mind-independent moral properties. Perhaps more importantly, we've learned that paradigmatically constructivist discourses like fashion are capable of supporting the very features (e.g., error, improvement, disagreement,

⁴⁹ On this point, see Railton 1986a, Boyd 1988, 181-7, and Brink 1984, 1989, 11-3.

predictive power, feedback mechanisms) that the Realists points to as evidence for the mind-independence of moral properties.⁵⁰ The upshot is that constructivism offers the best hope of vindicating the objectivity of ethics. The crucial question is which, if any, constructivist theory is able to pass the second test.

⁵⁰ I develop this point more in Kurth ms. 2.

Chapter 4: Constructivism and Moral Error

The arguments of chapters 2-3 entail that a viable Objectivist account—if there is one—will be some form of constructivism. As was the case in the last chapter, I propose to assess the plausibility of existing varieties of constructivism in virtue of their ability to provide a plausible account of moral error. And I will again use the second test to do this. The test, recall, demands that a viable Objectivist account must provide an account of moral properties and facts that shows how a moral judgment can be wrong, and do so in a way that is consistent with the content and credence conditions. There's reason to be optimistic. Unlike expressivism, constructivist accounts are varieties of descriptivism; so they can make use of truth-conditional accounts of meaning and inference. And unlike realism, constructivism grounds moral facts in something tangible and familiar—desires, affective states like anger and guilt, or certain deliberative capacities. So it seems better positioned to meet the credence condition. Moreover, most constructivist accounts mesh nicely with the commonsense thought that morality has both cognitive and conative/affective elements: we latch onto the world around us, including its moral features, both by reasoning about it and by reacting to it emotionally.¹

The project of investigating whether constructivists can meet the second test is complicated by the tremendous diversity of views that fall under the label. That said, looking at the various forms of constructivism allows us to distinguish three general

¹ See Nagel 1979 for insightful thoughts on this point.

families of views: those that look to distinctive sentiments or affective responses (sentimentalist accounts), those that appeal to an idealized or abstract vetting process (idealization accounts), and those that focus on certain rational capacities (rationalist accounts). I do not take there to be sharp lines between these three families. Nor, as we will see, do I suppose that they exhaust the logical space available to the constructivist. But the taxonomy proves useful for investigating some of the central features motivating constructivist accounts.

In what follows, we will see that existing constructivist views face a predicament much like the one that undermined realist accounts that appealed to explicit definitions: They confront a trade off between securing a reasonable account of the content of morality on the one hand, and providing a plausible account of the nature of moral facts on the other. But seeing why they face this dilemma will point us to better options.

1. Sentimentalist Views

By sentimentalist views I mean to pick out the family of accounts put forth by, for instance, John McDowell (1985), David Wiggins (1987), and—more recently—Shaun Nichols (2004) and Justin D’Arms & Daniel Jacobson (2006).² These views

² A few points of clarification and elaboration: (1) The label ‘sentimentalism’ is often used to pick out a distinctive realm of non-descriptivist views (Foot 1995 does this). But on my account, sentimentalism is neutral between descriptivism and non-descriptivism. The most plausible non-descriptivist versions of sentimentalism take the form of the sophisticated expressivism discussed (and dismissed) in chap. 2. The discussion that follows investigates the prospects of a descriptivist rendering of sentimentalism. (2) Among descriptivist sentimentalists we can distinguish between “old-school sentimentalists” (e.g., Shaftesbury, Hutcheson, Smith) and “neo-sentimentalists” (those listed in the text). The primary

share a common set of motivating assumptions. First, they take moral properties and facts to be constituted, in a sophisticated manner, by certain emotional or affective responses. The sentimentalist begins with the observation that there is an intimate connection between moral judgments about (say) wrongness on the one hand, and certain emotional or affective responses—feelings of anger or disapproval—on the other. They also note that our primary means of understanding what wrongness (say) is, is through the associated emotion or affective response (disapproval, anger). Though these features suggest that moral properties like wrongness are constituted by the associated responses, sentimentalists note that it would be implausible to take wrong acts to be *whatever* acts provoke anger or disapproval—the connection between property and response is more subtle. After all, one’s disapproval (or anger) can be misplaced because, say, you’ve misperceived the situation at hand. And even when disapproval (or anger) isn’t misplaced, it doesn’t always map to *moral* wrongs—it can be appropriate for me to disapprove of a fashion *faux pas* or an aesthetic blunder. So sentimentalists seek to specify particular sentimental responses that, when *appropriate* or *fitting*, are constitutive of the moral facts. Thus, we should take wrongness to be constituted by (say) *fitting* instances of anger.

This proposal provides the beginnings of a nuanced account of how one’s sentimental responses can go wrong. In particular, moral error is the upshot of responses that are, in the relevant sense, inappropriate or unfitting. But clearly this

difference between the two lies in the level of sophistication of the underlying account of the sentiments. The discussion that follows focuses on the more sophisticated “neo-sentimentalist” accounts.

strategy only succeeds in providing a plausible account of error if it can give a substantive explanation of what distinguishes merited responses from unmerited ones. It's not enough to be able to say that Janice has erred because her response wasn't merited. We need to be able to substantiate *why* it was unmerited—for without this there's no way to distinguish errant moral judgments from judgments one just happens not to dislike.

In the face of this challenge, these sentimental accounts offer a broadly teleological account of what counts as a fitting response: we distinguish the fitting from the unfitting instances of (say) anger in virtue of an understanding of what the *function* of feelings of anger is. Roughly, the idea is that our sentimental responses have distinctive (social-psychological, normative, etc.) functions. So by understanding what these functions are, we can identify the fitting responses—they're the ones that comport with the (teleological) function of the sentiment in question.³

To illustrate the sentimentalists' move to teleology, let's follow McDowell (1985) and focus on the non-moral case of fear—a case that he and other sentimentalist take to work in a way that is analogous to their proposals for good, right, and the like. To say that *x* is fearful is just to say it would be appropriate to fear

³ It is worth noting that one could be a sentimentalists about moral concepts, moral properties, or both. Considerations of ontological simplicity suggest that if one is inclined to a sentimentalist account of moral concepts, one should also be inclined to a sentimentalist account of the associated properties. However, while sentimentalism about both has been the standard route, some (e.g., D'Arms & Jacobson 2000, 2006) have recently sought to advance sentimentalism only about *concepts* (in the hope of being able to set aside thorny metaphysical questions). But given the nature of our investigation, I will presume that the accounts I am exploring are concerned to advance a sentimentalist account of both moral properties and concepts.

x. But since this statement explains the fearful in terms of appropriate feelings of fear, we have a circular account of what the function of fear is. So we can't make much progress in distinguishing between fitting and unfitting fears in a principled way. However, we can work our way into the circle in an *indirect* way. That is, we can get a sense of what makes fear appropriate by investigating the relation that holds between feelings of fear and, say, the presence of dangers and harms. Importantly, this does not result in a *definition* or an *analysis* of fear in terms of danger and harm; rather, it's an *explanation* or *gloss* of the function of fear. It's an explanation that allows us to demonstrate that our fear "responses are well-placed".⁴ Moreover, we get an independent (non-circular) method of determining what things are, in fact, fearful: rather than explaining the fearful in terms of appropriate instances of fear, we explain it in terms of the functional gloss we've developed. While sentimentalists typically grant that this move to teleology may not completely remove the circularity of their accounts, they deny that this is a problem. It's not a problem because they have provided (in the functional glosses) a substantive tool that allows us to understand (and distinguish between) fitting and unfitting responses.⁵

Even if we set the circularity issue aside, the move to a teleological account raises a significant metaphysical question—namely, what is the tie between a given response and the associated property such that the response is able to carve out the functional role that the associated property occupies? Is it a causal relation? A rational

⁴ McDowell (1985: 145-6). C.f., D'Arms & Jacobson (2006), Wiggins (1987: 194-5, esp. n16).

⁵ Wiggins (1987: 194-5, esp. n16) is most explicit on this point.

one? Something else? Without an answer to this question, we don't have an explanation for why we should take these responses to be *constitutive of*—rather than merely *evidence for*—the associated moral properties. One seemingly promising line of reply would be to offer an evolutionary/adaptive account.⁶ More specifically, we explain how the response carves out the associated functional role by giving an evolutionary account of the distinctive concern of the response in question. The general idea here is that certain basic emotions—things like anger, fear, envy, disgust, joy, sadness, surprise—are evolutionarily beneficial responses to features of the world that have particular significance for our survival. It's because these responses track things that tend to confer an adaptive advantage that they are—when fitting—constitutive of what's valuable. So by investigating what our responses have evolved to track, we can get some handle on the associated function. Consider: if the distinctive concern of envy is the loss of a significant positional good like status within one's group, then envy is merited when one has lost status.⁷

A benefit of this option is that it is widely accepted among psychologists, cognitive scientists, and others doing empirical work on emotions that there are basic emotions and that they are captured by something more or less like the above list. So we have a foundation for an account of moral error that meets the credence condition. However, this benefit also draws out the problem with this sentimentalist proposal: the move to basic emotions proves unworkable as the basis for a general account of

⁶ This strategy is employed in Wiggins 1987, Nichols 2004, and (with regard to certain non-moral values) D'Arms & Jacobson 2006.

⁷ An analysis of envy of this sort is offered in D'Arms & Jacobson 2006.

morality because it cannot meet the content condition. For notice, if one is to establish that basic emotions have *moral* import—that is, that they reflect more than just interesting psychological facts about us—one must provide some reason to think that an account of our moral discourse and practice can be built from facts about basic emotions. The trouble is that basic emotions just don't provide a particularly good base for explaining moral discourse. For one, though many have labeled some of the candidate basic emotions 'moral emotions' (e.g., regret, guilt, and anger), there still remains the substantial project of explicating the evolutionary tie between these basic emotions, and prototypically moral concepts and properties (e.g., right and wrong, good and bad). For instance, one must provide an account of the adaptive advantage that anger provides such that it maps in a plausible way to our concept of wrongness. Furthermore, other purported basic emotions such as amusement, surprise, and sorrow are not typically thought of as 'moral' emotions. So the advocate of the move to ground moral concepts and properties in basic emotions appears to face the awkward task of explaining—in a principled way—why *only some* of her esteemed basic emotions are relevant to morality. In short, it's very hard to see how a sufficiently broad account of the content of morality could be built using only basic emotions—it's just such a limited set of tools from which to build these explanations.

Might one be able to address this problem by going beyond the standard list of basic emotions? Perhaps, but the move faces substantial challenges that make success unlikely. For starters, expanding the list in order to capture a distinctly moral set of emotions would, without further explanation, be but an *ad hoc* maneuver to try to

satisfy the content condition. After all, the motivation of this strategy is to piggy-back on the credibility of the psychologists' evolutionary story. But if we go beyond their list, this credibility is lost.⁸

McDowell suggests an alternative possibility for explaining how we are to identify what the characteristic function of a response is—one based more in claims about rationality than evolution. He proposes that we understand the tie between a response and its functional role as the product of “relations of rational consequentiality” (1985: 146). Though McDowell does little to flesh this suggestion out, what he does say comes by way of extending the analogy with fear/danger discussed above. His proposal seems to be that by thinking about dangers and harms, we are able to take up a related, but distinct, vantage point on the function of fear—one not bound up in our *feelings* of fear, but rather our *thoughts* about dangers and harms. This alternative vantage point makes the rational consequence relations that hold between fear and its function “intelligible.” Fleshing this out, McDowell seems to suggest that for a given instance of our fear response, we can use a counterfactual test to determine whether the feeling is fitting and why. In particular, we specify some feature of the situation and ask whether we would be able to explain our feelings were we to deny the relevance of the feature in question. If we could not, then we can

⁸ In conversation, both D'Arms and Jacobson have indicated that it is for reasons like those expressed in the text that they endorse a sentimentalist account only for certain normative *concepts* and not for the underlying *properties*. They endorse an expressivistic account of the underlying “properties”. Though such a move may allow them to avoid some of the problems discussed above, it runs them straight into all the troubles discussed in chapter 2. Concerns of this sort have also inclined them to advance a sentimentalist account for only certain *non-moral* values—they're inclined to think that sentimentalism *cannot* explain morality.

conclude that the feature makes the feeling appropriate. We can conclude that “some of [fear’s] objects are really fearful, [and thereby] also make plain, case by case, what it is about them that makes them so” (146). Consider an example: What makes it appropriate to feel fear at the cliff’s edge? Presumably, it’s the fact that one could easily fall a long way. Moreover, the counterfactual test can confirm this: we would not be able to explain our fear response were it not the case that the situation was, in fact, dangerous (given the change of falling).

But we can grant McDowell’s point that this is not something we could “coherently deny” and insist nonetheless that it doesn’t deliver what he needs. To confirm his constructivism, McDowell needs to provide an account that demonstrates that our fear response is *constitutive* of the facts about what’s fearful. But the counterfactual test he provides is just too permissive. For notice, we have already seen (chap. 3) that moral functionalists use the same counterfactual test as an argument in favor of a *causal* account of a *mind-independent* moral reality!⁹ So, for all that’s been said, our sentiments merely play an epistemic role.

Stepping back, we’re now in a position to assess sentimentalism. The sentimentalist is right to stake out a substantive role for emotional and affective states in our understanding of moral properties and facts. He’s also right to think that this

⁹ Though I won’t pursue the point here, it is worth noting that McDowell also owes us an explanation of how one vantage point (danger) can make the rational consequence relations of another (fear) intelligible. McDowell is well aware that his explanation must avoid, as he says, “the bogus epistemology of intuitionism” (1987: 162). However, I doubt that his appeal to a “sensitivity for virtuous judgment” can do the needed work. See Jacobson 2005 for some insightful concerns about McDowell’s proposal.

role will be best understood via an understanding of the function that these emotional/affective states play in regulating our thoughts and actions. But as we've seen, he's wrong to maintain that a viable account of moral properties and facts—one capable of giving a plausible account of moral error—can be grounded in sentiments alone. The next two families of views—idealization and rationalist accounts—can be seen as attempts to add a cognitive or rational element to make up for the limitations of the sentimentalists story.¹⁰ Though I believe this strategy is on the right track, I also believe that these particular views ultimately fall short.

2. Idealization Views

Idealization accounts identify moral facts with the choices, desires, or inclinations one would have under certain idealized circumstances. This move taps into powerful intuitions: Facts about what's morally right or good are intimately connected to facts about what we would want, or would approve of, if we had all the relevant information and thought about it properly. It's ignorance and bad reasoning that trips us up. This intuitive resonance explains much of the appeal of idealization accounts. In their canonical form, these accounts identify rightness (say) with the choices that agents would make (or the things they would desire) were they to have full and vivid access to all the facts, and were they to assess those facts with perfect

¹⁰ I don't mean to imply that advocates of the idealization and rationalist views take themselves to be providing a way to patch up sentimentalism. Typically, they don't. The point is that looking to their proposals in this way can reveal something interesting about the nature of moral properties and facts.

reasoning.¹¹ The expectation is that by appealing to such idealizations (e.g., perfect reasoning, full information), one secures the separation from one's actual responses needed in order to make sense of error. But, as we will see, these views must allow that the idealizations involved could be extreme—so extreme in fact that they can't provide an account of error that meets the credence condition.

It's all too familiar that our epistemic and cognitive limitations prevent us from being able to make accurate choices about what we want, or what sort of rules we think are most appropriate. So though our actual deliberative capacities and beliefs can serve as a starting place for an account of moral facts, they cannot be the whole story. The appeal to idealization is made in order to address the epistemic and cognitive limitations that we, as we actually are, have. But given that these limitations are surely quite profound, the idealization involved will likely need to be severe—e.g., we must have *all* the relevant facts, we must have cognitive capacities that allow us to process these facts *flawlessly*. The resulting commitment for advocates of an idealization account, one that they typically acknowledge, holds that (i) an idealized agent can become (suitably) informed of a potentially infinite number of possible ways the world might be organized, including all the details of their costs and benefits, and that

¹¹ Views of this general form can be found in Firth 1952, Brandt 1979, and Smith 1995. It is worth mentioning that many have pursued a largely similar idealization strategy with regard to the *non-moral* good (e.g., Railton 1986b, Lewis 1989, Johnston 1989, Smith 1989). These accounts are subject to the versions of the criticisms that follow.

(ii) this agent can then make a choice about which he would want or would endorse—a choice that's constitutive of the moral facts.¹²

But both claims bring trouble. Reflection on claim (i) in light of general facts about human psychology and cognition draws out how radical the idealization process must be. Both empirical research and theoretical considerations demonstrate that how an individual is informed (e.g., how graphically the details are presented), and what state of mind he is in (e.g., sad, happy, depressed), can have significant impacts on the way that the information is processed. Thus the advocate of an idealization account must specify—in a principled manner—how the information is to be presented to the idealized agent, and what state of mind the idealized agent is to have. Because of the difficulty of providing such principled distinction, the best response to this concern seems to be to expose the idealized agent to the information in *all* its possible forms, and for *all* the possible states of mind he could be in. But, given the massive amount of information involved here, one might reasonably deny the plausibility of this move. Wouldn't being presented with so much information just overwhelm us?¹³

But the situation is worse for the claim about the choices that the idealized agent makes (claim (ii) above). These accounts suppose that the idealized agent is presented with the set of possible ways the world might be organized and then makes pair wise comparative judgments about which she would want or would endorse. But

¹² This is a bit of a gloss—the idealized agent is typically presented as not choosing what *he* would want, but what he would want his non-idealized self to want if he were in the shoes of that non-idealized self. For present purposes, these refinements need not concern us.

¹³ C.f., Rosati 1995, Velleman 1988.

clearly, if this process of making comparative judgments is to work, it must obey a transitivity principle, for transitivity ensures that comparisons across pairs are valid. That is, transitivity ensures that the agent can conclude that C would be desired over A from the comparative judgments that B would be desired over A, and C would be desired over B. However, the process of becoming informed about the various ways the world could be organized will likely affect the idealized agent's perspective about what she wants—as we've seen, *that such changes will occur is the whole point of idealizing*. But if the judging agent's perspective could—and likely will—change, then there's no reason to think that there will be a consistent perspective from which the idealized agent is making her judgments. The consequences here are significant, for we now see that there is no reason to think that the needed transitivity assumption will hold—it's too likely that the perspective from which the pairs $\langle A, B \rangle$ and $\langle B, C \rangle$ will be different.

There seem to be two ways to address this problem. First, one could seek to privilege a particular perspective for making the comparative judgments. Clearly, any choice would need to be justified. Without such justification one would not have an account of error—one would be unable to distinguish 'your judgment is wrong' from 'your judgment is one I don't like'. But given that we are dealing with such a highly idealized state of affairs, there just isn't a base from which to ground such a justification.¹⁴ Alternatively, one might try to push the idealization to a higher level of

¹⁴ Sobel (1994) raises a concern along these lines.

abstraction by assuming that the idealized agent makes her comparative judgments infinitely quickly—thus closing the gap in which one’s perspective could change.¹⁵ But if this move is to secure transitivity, it must *guarantee* the consistency of the idealized agent’s perspective. Achieving this requires that the temporal period over which the idealized agent makes her comparative judgments be vanishingly small—for otherwise changes in the idealized agent’s perspective will still be possible. But now we seem to have backed ourselves into a impossibility—namely, a temporal process (i.e., the idealized agent’s comparative judgments of $\langle A, B \rangle$, $\langle B, C \rangle$, ...) occurring timelessly. But any view that requires one to countenance such impossibilities surely fails the credence condition.

Finally, there are also very good reasons for doubting that the move to idealization could meet the content condition. As we’ve seen, the idealization process will likely require severe revisions to our psychological make-up in order to address the cognitive and epistemic deficiencies that we actually have. But given that such severe changes might be required, there is no guarantee that the judgments that result from this new, idealized perspective won’t clash violently with our commonsense judgments about what is good and bad, right and wrong. Why think that knowing all the facts, and reasoning perfectly about them, won’t lead to a world of grass counters as the most desirable option?

¹⁵ This options is hinted at by Railton (1986b: 57-8), and made explicitly on his behalf by Miller (2003: 215).

I take these problems to provide a decisive case against an account of moral facts that makes use of a radical form of idealization. But I do not take the discussion to here to undermine the thought that moral facts are the upshot of some cognitive vetting process. On this score, I think that we have much to learn from rationalist accounts.

3. Rationalist Views

The distinctive feature of rationalist accounts is the thought that morality is essentially concerned with reasons and reasoning. In what follows, I will focus on two general rationalist proposals. The first holds that moral facts fall out of basic facts about what it is to be an agent—to be a being capable of choosing its aims and goals. The second strategy takes moral facts to be the product of a hypothetical agreement among agents. Given the diversity that we find in both of these versions of rationalism, my strategy will be to explore some of the most worked out and influential versions of each. Looking at these dominant versions will allow us to identify a general set of concerns that tell against the agency and hypothetical agreement strategies.

(3.1) Agency Accounts. Christine Korsgaard's starting place in both *Sources of Normativity* (1996) and *Self-Constitution* (2009) is what she calls our "practical identities". Practical identities are socially shaped roles such as being a father, a Catholic, an outdoor-enthusiast, a neighbor. They not only give meaning and purpose to our lives, but also carry certain obligations. To take oneself to be a father *just is* to take oneself as having an obligation to care for one's children. Korsgaard's interest in

practical identities lies in her observations about the problems that they bring. For notice, each of us has many practical identities, a fact that entails that we also have different—often conflicting—sets of obligations. So we need a way of sorting out and prioritizing these various demands. But more importantly, Korsgaard emphasizes that we need a way of doing this that brings cohesion to our lives—a way that *unifies* us as the distinct and autonomous person we are (more on this below). The thrust of Korsgaard’s argument—and what I take to be the deep insight in her account—is that we can get a handle on the nature of morality by investigating what resolving these practical problems involves. The core of her argument goes like this:

- (P1) That I have various practical identities presents me with practical problems: I need ways of choosing between them when they conflict, and of unifying them into the coherent whole that is me.
- (P2) Though it’s a contingent matter *which* practical identities I have, that I have *some* practical identity or other is not: I must choose an identity in order to be an autonomous agent—to be a person in the normative sense.
- (P3) To choose an identity (i.e., to unify myself) is to confer value on that identity.
- (P4) To choose (and so to confer value) in this way requires a principle.
- (C1) Therefore, in virtue of the fact that I must choose an identity, I am committed to valuing of myself in a distinctive, principled way.
- (P5) But the principle by which I value myself (and the reasons it provides me) cannot be private.
- (C2) Therefore, the principles (reasons) by which I value myself must be the principles (reasons) by which I value others.
- (C3) Therefore, I am committed to valuing others in the very same way that I am committed to valuing myself.

If sound, Korsgaard's argument delivers (in (C3)) a familiar, substantive conception of morality—it's essentially Kant's Humanity Formulation of the Categorical Imperative. But there are a variety of reasons for thinking Korsgaard's argument falls short.

I want to begin with a general worry about the underlying account of persons that Korsgaard employs. She advocates for, and her argument depends on, a distinctly Kantian account of persons¹⁶—an account whereby one is a person just in case there is a principle that forms the basis of (constitutes) one's identity. On this account, a person is that which unifies a particular set of (embodied) inclinations and impulses; a person is something over and above these assorted impulses—namely, that which brings them together into a unified whole. In this way her account of persons departs from the more widely accepted psychological continuity accounts.¹⁷ So what pushes Korsgaard to wed her account of morality to a controversial thesis about the nature of persons?

Korsgaard's driving motivation is the belief that this account of persons best captures the connection between agency and responsibility. Korsgaard makes this point nicely in the following quotes:

It is because our actions are expressive of principles that we ourselves have chosen...that it makes sense for us to hold one another answerable for their actions. (2009: 131)

¹⁶ This may be too strong. Korsgaard 2009 argues that a comparable notion of person can be found in Plato.

¹⁷ Roughly, psychological continuity accounts hold that X is the same person as Y just in case X is (in the right way) psychologically continuous with Y. See, for instance, Parfit 1984 and Shoemaker 1984.

When you deliberately decide what sorts of effects you will bring into the world, you are also deliberately deciding what sort of cause you will be. And that means you are deciding who you are. So we are each faced with the task of constructing a peculiar, individual kind of identity—personal or practical identity.... It is this sort of identity that makes sense of our practice of holding people responsible... (19-20)

In short, the Kantian account of persons is plausible because of its ability to explain what it is about our choices that makes us responsible for them.

But grounding her account of persons in this way is troubling for a couple of reasons. First, it entails that we have a *moral* responsibility to choose and maintain a practical identity.¹⁸ That's odd. Intuitively, maintaining a practical identity is (at best) a personal ideal. But Korsgaard's account requires that we inflate this ideal to something of foundational moral importance. Worse, the proffered analysis is circular. Recall, the appeal of the Kantian account of persons is supposed to lie in its ability to explain the tie between agency and responsibility. That is, we're supposed to get a notion of persons suitable for attributions of responsibility. The explanation we get is that persons can be held responsible because they must choose a practical identity to live by. So far, so good. But when we ask *why* one must choose a practical identity, we're told that it's because we have a moral responsibility to make the choice. In short, persons are morally responsible for their choices because they have a moral responsibility to make those choices. Not only is this account uninformative given the platitudinous account of the tie between agency and responsibility, but it gives us no independent way to get into the circle of moral concepts that it employs. The upshot is

¹⁸ Korsgaard (2009: 174-5) is quite explicit about this.

that we have good reason to doubt the viability of the account of persons on which Korsgaard's moral theory rests.

I now want to turn to raise some more specific concerns about the argument leading up to the first conclusion (P1-C1). To begin, notice that (P3) presumes that one's choice functions to impart or create value. But, unless more is said, the mere fact that we choose an identity does not entail the constructivist claim that choice creates value. After all, one's choice might merely recognize value that exists independently of one's choice (a claim that is consistent, not with constructivism, but realism). Korsgaard doesn't think this is a problem because she thinks she can give us independent reasons for ruling out the realist option. These arguments maintain that realism fails because it can't capture the essential connection between willing an action, and taking up the means to achieve that action. Because of this, realism is committed to an "incoherent" account of agency (2009: 64-6; 1996: chap. 2). But notice this argument relies on a claim about judgment internalism. So it works only if judgment internalism is true. This is worth noting not only because the truth of judgment internalism is something many realists reject,¹⁹ but also because, as we'll see, it does not seem to be a claim that Korsgaard is in a position to make.

Setting this realism point aside for the moment, let's turn to investigate why Korsgaard thinks there *must* be a principle by which we choose our identities. First,

¹⁹ See, for instance, Brink 1989: chap. 3, Railton 1986a, Boyd 1988, and Bloomfield 2001.

notice that the necessity claim of (P4) is ambiguous between as strong and a weak reading:

Strong: There is a principle and we use it in order to choose our identities.

Weak: We choose identities and one could attribute the use of some principle to those choices.

Korsgaard is well aware that she needs the strong claim. Her defense of it comes from what I will call the Collapse Argument: When you make a choice about your practical identity, you are making a choice about who to be. But to make that kind of choice requires that you identify yourself with a principle. Why? Well, having such a principle makes it the case that this choice is “a product of your own identity”—that it is “expressive, or representative, of yourself” (2009: 75). Moreover, it *must* be this way. According to Korsgaard, were you to be capable of choosing *without* a principle, you would lose the ability to distinguish yourself from the desires and inclinations from which you act. The result would be the collapse of the person you are supposed to be: choice without a principle would entail that “you are not one person, but a series, a *mere heap*, of unrelated impulses” (76, emphasis original).²⁰

But the Collapse Argument is problematic for a variety of reasons. For starters, there are reasons to think that collapse claim is just false. Do I really cease to be an agent (or lose my reason to live) if there is no principle that undergirds the manner in which I negotiate the conflicts among my practical identities? No. So long as I could

²⁰ A similar argument can be found in Korsgaard 1996: 120-2—though there the collapse is put more starkly: “unless you are committed [by way of a principle] to some conception of your practical identity, you will lose your grip on yourself as having any reason to do one thing rather than another – and with it, your grip on yourself *as having any reason to live...*” (120-1, emphasis added).

retrospectively discern some principle in the choices that I have made, I would be able to mark a distinction between the person that is me, and the impulses from which I have acted. So I wouldn't collapse into a "mere heap" of impulses. That is, Korsgaard overstates her case—she fails to notice that a "collapse" can be avoided given only the *weak* version of the necessity claim.

These considerations suggest that we might be reading the collapse claim too literally. I don't *literally* cease to be an agent if I fail to act from principle; but my agency is seriously compromised nonetheless—I fail to live up to a significant regulative ideal. However, if we go with this less literal reading, then it's hard to see how the resulting conclusion about the necessity of a principle can serve as more than a guiding principle or a rule of thumb. This weaker conclusion is problematic. For starters, Korsgaard aims for something stronger—a constitutive claim about what it is to be an agent. So accepting the weaker conclusion would be a concession. Moreover, given her critical arguments against realism, she *needs* something stronger. Here's why: If acting from principle is merely a regulative ideal and not a constitutive feature of what it is to be an agent, then there is nothing *incoherent* in failing to will the means to one's ends (though there is something problematic). But if there's nothing incoherent here, then the case against the realism falls apart and with it the justification for (P3).

In light of all the problems with the strong necessity claim, why is Korsgaard inclined to think it must be true? What's her beef with the weak version? She rejects

the weak version because she thinks that it fails to give a plausible account of what it is to choose or decide. Witness:

when you determine yourself to be the cause of the movements which constitute your action, you *must* identify yourself with the principle of choice on which you act. For instance, suppose you experience a conflict of desire: you have a desire to do both A and B, and they are incompatible. You have some principle that favors A over B, so you exercise this principle, and you choose to do A. In this kind of case, you do not regard yourself as a mere passive spectator to the battle between A and B. You regard the choice as yours, as the product of your own activity because you regard the principle of choice as expressive, or representative of yourself—of your own causality. (75, emphasis added)

According to Korsgaard, the reason why we can't explain choices like this in terms of some third psychological feature is that it commits what we could call, for reasons that will become apparent, the pineal gland fallacy (99-100). It's not sufficient to just point to some fancy psychological state—the proverbial pineal gland—as an answer to how we're able to make the choices that we do. That's just to rename the original process we were trying to explain. What's needed is an explanation of how that state makes choice possible. Without this last step, you're guilty of the pineal gland fallacy. The concern is legitimate. But notice—as an argument against the weak version of the necessity claim—it merely serves as a demand for an explanation. Moreover, as will become apparent (chap. 6), I think this demand can be met. In fact, some of Korsgaard's own observations—principally her observation that we must address the conflict that exists among our practical identities—will help us see how.

In addition to the problems with Korsgaard's first argument, there are further problems with the argument that aims to demonstrate that, because the reasons that I

have to value myself are public reasons, they are reasons to value not just myself, but anyone else (C1-C3).²¹ The key to this argument is premise (P5), which holds that there can be no private reasons. Korsgaard defends this premise in chapter 9 of *Self-Constitution*.²² This argument rests on the claim that the possibility of collective *action* requires that collective *deliberation* be possible. But if there's to be collective deliberation, as surely there is, then there must be public reasons—there must be reasons that “extend across the boundaries between persons” (191). But this argument is problematic for a variety of reasons. First, even if public reasons are essential for *collective* deliberation and action, that fact is not sufficient to show that private reasons aren't employed in *individual* choice and action. But this means that Korsgaard's (C2) can't be true—for it holds that the principles that give me reason to value myself needn't *also* be reasons for me to value others. Second, it is not clear that collective choice requires public reasons—at least if are starting place is a non-question begging account of collective action. That is, if we understand collective action as multiple agents collaborating to achieve some end, then it seem possible that

²¹ Since I take the above arguments as sufficient to undermine the plausibility of Korsgaard's account, the discussion of this second part of her argument will be brief. Though I think that there are ways that Korsgaard could respond to the concerns noted below, I think they ultimately fail. I hope to give this argument a more though treatment in later work.

²² A couple of points: (1) In the *Sources of Normativity*, Korsgaard presents what—at least on the surface—appears to be a different argument against the possibility of private reasons. There she argues that for the reasons *analogous* to those given by Wittgenstein to shows us that there can be no private languages, we can see that there can be no private reasons. But since I think Wittgenstein's argument fails for reasons others have pointed out (e.g., Soames 2005), I think Korsgaard's must as well. (2) It's also worth mentioning that Korsgaard notes that the private language argument of *Sources* has been heavily criticized. It is not clear however whether she herself takes those criticisms to be decisive. The fact that she offers a new argument suggests that she is not entirely happy with her original case against private reasons.

this could happen without public reasons. What *would* be necessary is that agents have sufficient similarity among their private reasons.²³

Summing up, what general conclusions does our investigation of Korsgaard's account allow us to draw about the prospects for other agency accounts (e.g., Kant 1785, Nagel 1970, Gerwitz 1980)? First, we see that any account of agency capable of providing a foundation for moral facts will almost surely—if it's to avoid circularity—need to rely on an implausibly strong account of persons or agency. But even if this problem is surmountable, there remains the daunting task of explaining why the reasons and principles that govern an agent's *own* choices also serve as reasons and principles governing his choices concerning *others*. Concerns of this sort have pushed some rationalists way from the agency model and toward one focused on hypothetical agreement.²⁴

(3.2) Hypothetical Agreement Accounts. The unifying theme of hypothetical agreement accounts is that they see the content and normativity of morality as a consequence of some hypothetical agreement. The moral rules (facts) just are the obligations, prohibitions, etc. that we would agree to under certain hypothetical circumstances. And these rules have normative force for us in virtue of the fact that we would agree to them. Clearly, the plausibility of these accounts is tied to the plausibility of the hypothetical agreement process that they propose. Though there is

²³ A final worry: As Korsgaard herself admits, public reasons are not things of the empirical world, but the noumenal world (2009: 174-5). Though we might ultimately have to accept such a conclusion, there is strong *prima facie* reason to take this move to the noumenal realm to violate the credence condition.

²⁴ See, for instance, Scanlon 1998: 5-6 and Gauthier 1985: chap. 1.

much diversity in the details, we can discern two general strategies—those that propose a moralized agreement process, and those that don't. This partitioning of the accounts reveals that both face a substantial challenge. Moralized conceptions seem troublingly circular—the agreement process does not *construct* moral facts since the underlying conditions of the agreement already *build in* the moral substance. Non-moralized accounts face the opposite problem: convincing us that the constraints that they construct are—ones with moral upshot—ones we (morally) ought to follow. The discussion that follows looks at an influential version of each: T.M. Scanlon's moralized conception and David Gauthier's non-moralized one. What makes these accounts interesting is that they explicitly take on their respective challenges. Moreover, seeing why they nonetheless prove unable to address them will provide us with good reason for thinking alternative versions of each strategy are unlikely to succeed.

(3.2.1) Scanlon and Moralized Agreement. In *What We Owe to Each Other* (1998), Scanlon defends an account of reasons that is based on a general non-natural realism about value (96-8). But he maintains that these general facts about *value* are insufficient on their own to give us an account of *moral* facts—particularly, facts about right and wrong. These moral facts, Scanlon argues, are constituted by a contractualist-style hypothetical agreement procedure. Scanlon's insight lies in developing a procedure that gives center stage to the observation that morality is intimately concerned with justifiability to others. This concern with justification, after all, is a deep, driving feature of our moral thinking. The account that results holds that

facts about right and wrong are facts about which principles we would endorse given (i) that we have reason to live with others on terms that they could not reasonable reject, and (ii) our judgments about what would count as reasonable grounds for rejecting a given principle.

As noted above, Scanlon's account is a moralized one. We're now in a position to see why this is, and why it threatens to make his proposal viciously circular. The account is moralized because it analyzes wrongness in terms of principles that we judge to be reasonably rejectable; but what makes a principle reasonably rejectable is that it we would judge it to have the relevant wrong-making features (213-4). So, for example, we're to understand the *wrongness* of murder in terms our judgment about the *wrongness* of certain harms as providing reasonable grounds for rejection. And now we can see why it's reasonable to worry about the circular nature of Scanlon's proposal: For one, the account it provides seems trivial—e.g., the reason why it's wrong to murder is that it's wrong to inflict the harms that murders cause. Furthermore, because the account requires us to have an antecedent understanding of, for instance, the wrongness of certain harms in order to understand the wrongness of murder, it leaves with no independent way of making sense of what moral wrongness (in general) is.

Scanlon acknowledges that his account is circular. But he denies that it is viciously so:

While it would be objectionably circular to make “reasonable rejection” turn on presumed entitlements of the very sort that the principle in question is supposed to establish, it is misleading to suggest that when we are assessing the “reasonable rejectability” of a principle we must,

or even can, set aside assumption about other rights or entitlements all together. (214, emphasis added)

As the quote suggest, Scanlon backs up his denial that the circular nature of his account is problematic by maintaining that it is *impossible* to substantiate one moral judgment without appealing to other moral judgments. This is a point he emphasizes when he claims that “a sensible contractualism...will involve a holism about moral justification” (214).²⁵ Here the obvious worry is that holistic moral justification can support monstrous moral systems—systems that clearly violate the content condition. Scanlon denies that this will happen because he maintains that his proposal will deliver “definite conclusions about the grounds for reasonable rejection” that allow us to draw substantive conclusions about right and wrong that comport with commonsense morality (218). For instance, he argues that his contractualist procedure entails that it will never be permissible to break a promise in order to avoid a minor inconvenience (298-300). But there are at least two reasons to be unsatisfied with this response. First, until we’re told what counts as a “minor inconvenience,” there’s no reason to think that Scanlon has managed to move from trivialities to claims with substantive content. Granted, Scanlon gives a set of examples of considerations that *would* warrant breaking a promise (e.g., situations involving emergencies and threats).

²⁵ In light of comments Scanlon makes elsewhere (96-8), we can see that the argument here is just a version of Moore’s (in)famous Open Question Argument. But given the consensus that Moore’s argument, at best, causes problems for views that seek an *analytic* or *definitional* reduction of the moral to the non-moral, it’s hard to see why Scanlon thinks his claim is well founded. (For discussion of Moore’s argument, see, for instance, Frankena 1939, Brink 1989: 151-4, 162-3, Darwall, Gibbard & Railton 1992: 115-20, and Soames (2005: 34-70). I sketch my own concerns with Moore’s argument in Kurth 2007).

But he also insists that the list must be open ended, as guided by the “‘standard’ reasons’ that count against promise breaking” (299). So there is good reason to doubt that we’ve actually made much progress.

But there’s a more significant problem: when we look to see what substantiates Scanlon’s confidence that his account will deliver substantive conclusions about right and wrong, we see that it’s not the contractualism that does the work, but rather the underlying value realism. He is certain that it’s reasonable to reject a principle that would permit breaking promises in order to avoid minor inconveniences because of the *value* that our promises have to those we promise to do things for—where this value is understood on a realist, not a constructivist model (298-9). The upshot is that Scanlon’s response to the circularity charge “works” only by shifting the substantive work from the contractualist proposal, to his underlying non-naturalist realism about value. That is, he solves the circularity problem that his account faces by trading his constructivism for realism.²⁶ Moreover, there seems no reason to think that other moralized hypothetical agreement accounts won’t fall to a similar dilemma: They can’t avoid vicious circularity without giving up the very thing that makes their proposals distinctive and interesting—namely, the claim that moral facts are constituted by a hypothetical agreement.

(3.2.2) Gauthier and Non-Moralized Agreement. In *Morals by Agreement*

David Gauthier defends a non-moralized hypothetical agreement account. According

²⁶ Given the problems that we have uncovered for non-naturalist accounts (chap. 3.3.3), this does not amount to progress.

to Gauthier, moral rules are to be identified with the set of impartial constraints that purely self-interested agents would agree to in a hypothetical bargaining situation. While Gauthier grants that these constraints may not correspond exactly with commonsense morality, he believes that they will be significant overlap—significant enough to be tolerably revisionary (6).²⁷ Moreover, Gauthier maintains that this conclusion can be demonstrated from “a weak and widely accepted conception of practical rationality,” namely that given by rational choice theory (17). In this way, Gauthier’s proposal offers a sharp contrast with the proposals we’ve looked at from Korsgaard and Scanlon; we have a proposal that relies on neither a robust notion of agency/persons, nor a moralized foundation. That’s significant.

Though there is much that is interesting in Gauthier’s account, for present purposes, I want to focus on just one of its central features—his explanation for why we, here and now, are obligated to obey whatever a set of rules that is derived from a hypothetical agreement. According to Gauthier, an obligation to be moral

does not emerge like a rabbit from the empty hat. Rather,...it emerges quite simply from the application of the maximizing conception of rationality to certain [prisoners’ dilemma-style coordination problems]. Agreed mutual constraint is the rational response to these [coordination problems]. (9)

Thus, we, here and now, are obligated to be moral because it’s possible to show, given just the (weak) tenets of rational choice theory, that accepting the rules and constraints

²⁷ While there is reason to question whether the impartial constraints that come out of Gauthier’s hypothetical agreement process will actually meet the content constraint, I will not pursue that issue here. See, the discussion of chap. 5.2 for reasons for thinking Gauthier’s claim is plausible.

of morality is in our rational self-interest. We have, that is, a rational requirement to adopt the rules and constraints of morality.

However, Gauthier's response proves to be unable of doing the needed work. To see why, we need to draw out two key features of his argument. First, there is his commitment to a thoroughgoing Humean account of value and motivation. This commitment follows from his endorsement of an austere starting place and, in particular, his appeal to the rational choice methodology that understands rationality in terms of the maximization of individual preference (value). Here's why: Given Gauthier's purposes, his model of rational choice must have normative force in the sense that it deems some of our choices rational, others not. This in turn requires that we be able to distinguish our preferences from our choices. After all, if what you prefer just is what you choose, then your choices will *always* be rational. So rational choice theory will be unable act as a normative constraint.²⁸ To avoid this trivialization of claims of rationality, Gauthier moves to identify value with preferences, and equates choice with behaviors—e.g., I can be said to have chosen the apple if I reach for it *and* verbally confirm this as my preference with something like, "I'd like an apple" (27-8). The upshot of this is a Humean conception of value and normative/moral motivation: "[w]hat is good is good ultimately because it is preferred, and it is good from the standpoint of those and only those who prefer it" (59). The second feature to note is Gauthier's commitment to a Kantian rationalism about the

²⁸ This is a bit of a gloss, but it will suffice for present purposes.

bindingness of morality. According to Gauthier, moral constraints have force for us regardless of our desires and affections—morality must be, and can only be, rationally binding (103).

Now here's the problem: the combination of these Humean and Kantian commitments entails that the reasons generated by Gauthier's rational choice model have force for us only insofar as we *want* to be moral: The rational choice model tells us that it's rational to be moral, and the Kantian thesis delivers this as an essentially rational demand. But unless we already have a preference for being moral, we will—by the Humean thesis—have no interest in what the model demands we do. So while Gauthier's account may be able to confirm that our desire to be moral is rational, it seems unable to convince us to be rational if we're not inclined that way.

Now Gauthier has a ready response. His rational choice model can demonstrate that *anyone*—including those uninterested in being moral—has purely self-interested reasons to want to be moral. This is because the model shows us that morality will “make possible the more effective realization of one's interests, the greater fulfillment of one's preferences, whatever one's interests and preferences may be” (103). However, while it may be true that a purely self-interested person would find this response compelling, we—as we actually are—are not the purely self-interested agents of Gauthier's hypothetical bargaining situation. For one, we are not driven by purely self-interest. But, more importantly, it's reasonable to think that many of us would be repulsed by the prospect of adopting the reasoning of a purely self-interested agent—

we just don't want to be that kind of person. So there's no reason to think that this argument will have force for (many of) us.

At this point one might protest that I'm demanding too much in requiring Gauthier to show that actual individuals like us will be willing to accept reasons meant for purely self-interested agents. But this is a success condition that he does (and must) set for himself. Because he's committed to such an austere starting place, he holds that we must "begin from an initial presumption against morality" (8). So the burden of proof is on him. He must be able to provide a compelling case that we as we actually are have reason to be moral. But given the above, we can see that the burden has not been met.

4. The Verdict on Constructivism

We've discovered that constructivist accounts that appeal to robust assumptions about the nature of our cognitive, conative, and affective capacities seem incapable of providing an account of moral facts that pass the credence condition. More specifically, we've seen that in order to secure a suitable account of the content of morality, existing constructivist accounts need to rely on unprincipled, circular, or down-right implausible assumptions about our psychological and deliberative capacities. But given the tremendous diversity of constructivist accounts, we cannot say that there is *no* version of constructivism capable of passing the second test. In fact, as we will see in the chapters that follow, combining insights from the critical investigations of the realist and constructivist efforts to provide a plausible account of moral error point to an un(der)explored part of logical space available to descriptivists.

But we can start by culling some lessons from chapters 2-4 regarding what a viable account of the objectivity of ethics will look like:

An empirically tractable form of constructivism. The inability of expressivism to provide a plausible account of logic and inference led us to descriptivism. Yet we found that realists were unable to substantiate moral properties as mind-independent entities. This pointed us to constructivism. But as the critiques of McDowell, Korsgaard, and Scanlon revealed, constructivists fare better to the extent that they can ground their accounts in an empirically tractable manner. On this point at least, sentimentalism seems to be on the right track.

A moderate cognitive/rational element. Moral facts are the upshot of more than just certain emotional or affective responses—there also needs to be a cognitive or rational element. But this element must be modest. The investigation of idealization accounts reveals that radical forms of idealization are—at best—of dubious coherence. While this doesn't completely rule out the use of idealization, it does limit the degree to which it can be plausibly applied. Furthermore, our look to rationalism demonstrates how easy it is to slide from plausible claims about a role for reasons, deliberation, and justification to a set of implausibly strong assumptions, or an unhelpfully circular analysis.

A more deflated conception of moral properties & facts. The first two lessons result, at least in part, from difficulties that realist, idealization, and rationalist views face in providing an account of moral properties that could support a robust notion of fallibility. But as the discussion of the role functionalists' analogy argument demonstrates (chap 3.4), even paradigmatically constructivist discourses (e.g., fashion) support a rich set of critical practices. This suggests that many descriptivists err in seeking to make too much of the objectivity of ethics. Moral properties are likely best understood as being more akin to the properties of etiquette or fashion, than to the properties of physics and chemistry.²⁹

²⁹ This observation, when combined with the account that I will develop in the chapters 5-6 suggests that morality (and likely other areas of normative discourse) involves a novel type of property. Moral properties are not natural kinds like gold in the sense of being the upshot of something mind-independent, nor are they phenomenological/psychological kinds like (plausibly) pain in the sense of being wholly the upshot of certain internal psychological processes. Rather, moral properties are what we might call a social kind; they are the product of a distinctive set of social-psychological

Taking actual moral judgments as a starting place. The need for an account that is both metaphysically deflated and empirically tractable suggests that taking a closer look at how we actually go about making moral judgments may point us to a better understanding of the sorts of thing moral properties and facts must be.

These lessons suggest that a better way of vindicating the objectivity of ethics might lie in a form of moral conventionalism—the view that takes moral fact to be fixed (in some way) by social conventions. Let's turn to consider what such an account might be to say about the objectivity of ethics.

mechanisms. I hope to explore this idea, and develop an account of the underlying metaphysics, in future work.

Chapter 5: Transforming Moral Conventionalism

Suppose one is skeptical about the metaphysical, epistemological, and psychological claims on which realist, rationalist, ideal observer, and other robustly “objectivist” accounts of moral facts rest. Suppose further that one is wary of the deep revisionism required by error theories, non-cognitivism, and expressivism. That is, suppose you’ve found the arguments of chapters 1-4 compelling. Then there’s much to like in the conventionalist accounts offered by David Hume, Gilbert Harman, David Wong, and others.¹ By taking moral facts to be a function of social conventions, these accounts appear well positioned to make sense of the content and normativity that we take morality to have. And they seem to do so in a way that’s both naturalistically acceptable and metaphysically conservative.

Furthermore, because they take moral facts to be determined by our social conventions, conventionalists seem well positioned to provide a plausible account of moral error: a particular moral judgment is incorrect just in case it is inconsistent with the relevant convention. In fact, the ability to locate a source of moral error outside an individual’s own moral judgments seems to give conventionalists a significant advantage over their subjectivist rivals.² Though the subjectivist can also provide a

¹ Though Harman and Wong advertise themselves as relativists not conventionalists, their accounts of moral facts give an essential role to conventions (see, for instance Harman & Thomson (1996: chap. 2) and Wong (2006: 37-43)). Because of this, it is appropriate to classify their accounts as conventionalist accounts. I will say more about this in the next section.

² While the label ‘subjectivism’ is sometimes taken to include non-cognitivist accounts (e.g., Foot 1995), in this chapter and the next, I use it to refer to a cognitivist meta-ethic—one that takes moral

naturalistically acceptable and metaphysically conservative account of the normativity of ethics, his move to take moral facts to be fixed by individual judgments has been traditionally thought to prevent him from giving a plausible account of moral error.³ In light of all this, it would seem that conventionalism offers a particularly compelling account of our moral discourse and practice.

I disagree. In what follows, I will argue that one will be better able to secure what seems distinctive and interesting about conventionalism by transforming the standard conventionalist account into a sophisticated variety of moral subjectivism. I begin by demonstrating why many have found conventionalism a compelling meta-ethical view—as we will see, when combined with a plausible moral psychology, it offers a powerful and insightful explanation of our moral discourse. But despite this potential, conventionalism proves untenable: initial appearances to the contrary, its explanation of moral error is deeply problematic. At the end of the chapter, I will suggest that if we move from moral conventionalism to a sophisticated variety of moral subjectivism, we're likely to get a more plausible account of error, while retaining the core of what's distinctive and appealing in moral conventionalism. Developing this insight will be the work of chapter 6.

properties and facts to be constituted, in some way, by the subjective judgments of individuals. I will also say more about the relations between subjectivism and relativism in the next section.

³ Witness Gilbert Harman: “We do not ordinarily suppose that right and wrong are determined by a particular individual’s decisions and principles [as the subjectivist maintains]. We are inclined to think that morality has an external source, not an internal one” (1977: 92).

1. Moral Conventionalism—First Steps

(1.1) Conventions and Conventionalism. Let's begin by distinguishing between conventions and conventionalism. A convention, moral or otherwise, is a self-enforcing regularity in behavior that a group of individuals adopts, explicitly or implicitly, in order to reap the benefits of coordinated behavior.⁴ Paradigm examples of conventions are things like our use of bills and coins as mediums of exchange, our use of 'dog' to refer to a particular type of animal, and our practice of driving on the right side of the road. Conventionalism by contrast is the philosophical view that the nature and content of some area of discourse—money, language, rules of the road, morality—is to be (wholly) explained in terms of the conventions of that discourse.

Consider a paradigmatic example of conventionalism: money. Our use of bills as a medium of exchange is a convention that arose to address a distinctive, recurring coordination problem—namely, how to facilitate the exchange of goods and services among a diverse group of individual buyers and sellers. Now there are a variety of ways that a group could solve this coordination problem in the sense that a group could adopt any of a number of things (paper notes, coins, cigarettes, shells...) as their medium of exchange. Though individual members of the group might have different preferences about how to solve the coordination problem (some might prefer coins, others shells), they all prefer that there be *some* solution to no solution at all. But once

⁴ For the purposes of this essay, I presume that this broadly Lewisian account of conventions is correct. For an elaboration and defense, see Lewis 1969, Sugden 1986/2001. For critical alternatives, see Gilbert 1989 and Millikan 2005.

a group of individuals converges on a particular solution (the use of bills, say), a convention is fixed. More specifically, to say that a convention is fixed is to say that the group in question has adopted a self-enforcing pattern of behavior to address their coordination problem. So we should see a convention as articulating a distinctive conditional preference: each individual in a group G chooses to adopt a particular solution S to a recurring problem given that the others in G do so as well.

The money example allows us to highlight several distinctive features of a conventionalist account of a discourse. First, the conventionalist maintains that once we have an account of the conventions underlying the discourse in question, we have a complete account of the nature and content of that discourse. Second, a given area of discourse is distinguished by the distinctive coordination problems that it involves (e.g., how to facilitate the exchange of goods).⁵ Let's now turn to how these features get fleshed out by the moral conventionalist.

(1.2) Moral Conventionalism. Moral conventionalism is the meta-ethical view that seeks to explain our moral discourse and practice (entirely) in terms of social conventions.⁶ To help draw out what's distinctive in moral conventionalism, contrast it with realism. Both a realist and a conventionalist will agree that moral discourse employs a set of conventions. But they will differ over what that role is. For the

⁵ Some, e.g., Harman (Harman & Thomson 1997: chap. 8), also point to distinct methods of enforcement as a further dimension of differentiation.

⁶ We can distinguish between pure and hybrid version of moral conventionalism: while pure forms maintain that the entirety of our moral discourse can be explained in terms of conventions (e.g., Harman 1977), hybrid versions hold that conventions alone are not enough (e.g., Hume's move to add sympathy to his conventionalism). In what follows, I will focus primarily on pure versions, though much of what I say will carry over to hybrid accounts.

conventionalist, the conventions are—in some way—determinate of the nature and content of morality; for the realist, the conventions at best track a set of mind-independent (and, *a fortiori*, convention-independent) properties and facts.

As we've seen, we can also distinguish moral conventionalism from other varieties of conventionalism (e.g., conventionalism about etiquette and other social practices) by looking to the distinctive set of coordination problems that it takes moral conventions to have evolved to address.⁷ Though the moral conventionalist might specify what's distinctive of moral coordination problems in a variety of ways, I want to pursue what I take to be the most promising strategy—one that appeals to an intuitive account of moral psychology. Elements of the psychological account that follows are most prominent in the work of David Wong (1984, 2006); but they are also evident, at least to some extent, in David Hume (1888) and Gilbert Harman (1975, 1977, Harman & Thomson 1996). As we will see, this strategy significantly enhances the power and plausibility of the conventionalist's account. (It will also form a core part of the subjectivist proposal presented in the next chapter.)

The foundation for this conventionalist's moral psychology lies in two deep and familiar facts about human psychology.⁸ First, we are beings with diverse, often complex, personal projects. Our personal projects include things like nearer-term aims

⁷ There are several reasons why it's important that the moral conventionalist be able to make this distinction. For one, we want a way of capturing the intuitive difference between bad moral behavior and (say) bad manners. But more importantly, without an account of this intuitive difference, the moral conventionalist lacks a way to justify our practice of treating moral failures more seriously than bad manners.

⁸ See, for instance, Wong 2006: chap. 2 and Harman 1977: chaps. 8-9.

and goals—completing a book chapter, participating in a neighborhood clean up—as well as longer-term projects—raising a family, developing a particular career. Personal projects are psychologically significant because developing and pursuing them gives distinctive meaning to our lives. It gives us a distinctive sense of accomplishment and purpose—one that mere wishes and fantasies cannot provide.⁹ Second, we are social beings: our lives depend on, and are shaped by, social connections and social institutions. We rely on friends, family, associations, and the like for support and guidance, and these connections influence the values and goals that we develop. In these and other ways, social connections and institutions also give distinctive meaning to our lives.

But as is all too familiar, personal projects can conflict—my goal of advancing my political career may be incompatible with my wife’s desire for a quiet family life. Given the nature of our psychology, these conflicts can be destabilizing. For instance, when one’s own personal projects conflict with those of others, one’s ability to pursue one’s personal projects may be compromised.¹⁰ Given the connection between one’s

⁹ A couple of points: (1) Personal projects attach (in various way) to the basic human needs and interests that contribute to our well-being. (2) My account of personal projects has much in common with Korsgaard’s (1996, 2009) account of “practical identities” (see chap. 4.3.1), and Seana Shiffrin’s discussion of “complex ends” (1999: 784-8). Issues about the nature and significance of personal projects is something I hope to pursue in further work.

¹⁰ There are a variety of ways in which such conflict could compromise your ability to pursue your own personal projects. For one, as is often the case between intimates, one of your personal projects might be to help another realize her own personal projects. But, as in the husband and wife case in the text, this project might conflict with other projects that you have. Alternatively, given that we are social beings, we are troubled when we disrupt social cohesion. Because of this psychological fact about us, we’re likely to take preserving social cohesion as a personal project. So when we see that our own personal projects frustrate the personal projects of others (intimates or not), this project will be compromised.

pursuit of one's personal projects, and the meaning these projects give to one's life, the psychological toll exacted when they are frustrated can be significant. In fact, the more significance that a threatened personal project has for an individual, the more likely it is that that individual would suffer substantially were the project frustrated. But because particular social institutions (e.g., family, community) typically become invested in the individuals whose personal projects they help shape and support, conflicting personal projects can also threaten social cohesion. For instance, the cohesion of a community that comes together to support the personal projects of a particular individual may unravel when that individual's projects are frustrated.¹¹ But noting the significance that our personal projects and social connections have for us suggests that we have distinctive psychological needs: We have a psychological need for stable mechanisms that (i) allow us to assess and balance competing personal projects, and that (ii) facilitate the resolution of conflicting personal projects in a way that preserves and promotes social cohesion.

Given this intuitively plausible psychological picture, the conventionalist can explain which coordination problems are distinctive of morality: they are those that concern conflicting personal projects that have the potential to disrupt social cohesion.¹² To begin to see the virtues of this account, notice that it provides the conventionalist with a principled way of distinguishing moral conventions from non-

¹¹ Witness the disarray that erupted among the community of supporters for John McCain's 2008 Presidential bid once his campaign began to fail.

¹² See, for instance, Wong 2006: 37 and Harman 1977: 105. This general functional rendering of the moral is endorsed by both non-conventionalist moral philosophers (e.g., Korsgaard 2009, Gauthier 1985) and social scientists doing work on morality (e.g., Haidt & Kesebir 2009, Frank 1988).

moral ones (e.g., the conventions of etiquette or fashion). The conventionalist can demarcate moral and non-moral conventions in terms of their distinctive functional roles. He can argue that the function of morality is that of providing stable structures that allow for a principled way of assessing and balancing competing personal projects.¹³ The job of moral principles, moral reasoning, and the like is that of helping us adjudicate conflicting personal projects in a way that preserves social cohesion.¹⁴ By contrast, the conventions of (say) etiquette serve a different function—that of facilitating social interaction.¹⁵

Though I will discuss the benefits and costs of this conventionalist proposal in the following sections, it makes sense to take up one concerning the scope of the conventionalist's account here. On the one hand, one might worry that the focus on coordination problems leads to an account of morality that is too *narrow*. While morality is certainly concerned with coordinating behavior in ways that are mutually advantageous, it also seems concerned with matters like beneficence and justice where there is no (obvious) overlap of the interests of the involved parties. On the other hand, one might worry that this scope of this conventionalist proposal is too *broad*. Because

¹³ See, for instance, Wong 2006: 39 and Harman 1977: 91-6.

¹⁴ Does the function of morality also concern the working out of intra-personal conflicts? Perhaps, but since this is more controversial, and since nothing in the discussion that follows turns on it, I will not take it up. C.f., Wong 2006.

¹⁵ There may be no way of sharply delimitating the functions of various conventions—moral, etiquette, general prudence, and the like. If so, the conventionalist will not be able to give a precise account of the moral or other social domains. While one might take this as a reason for concern, I think so doing would be a mistake. For starters, there is the challenge of expressing this concern in a way that does not beg the question against the conventionalist. Moreover, what reason is there to think that there are sharp lines here? Are grotesque table manners morally offensive? If the behavior causes substantial harm or discomfort, perhaps. But given how mushy our intuitions are here, it seems reasonable to take cases like this as evidence of a vague boundary between morality and etiquette.

the conventionalist framework alone places very few constraints on the possible ways of solving a given coordination problem, it seems to lack the resources for an adequate account of the content of morality (recall the diversity of objects that have been accepted as mediums of exchange). We do not want an account of the content of morality that is radically at odds with our commonsense conception.

I believe that the conventionalist's psychological account points to a response to these concerns. Recall that the conventionalist's psychological assumptions suggest that morality serves a specific function—namely, providing stable mechanisms for the assessment and adjudication of competing personal projects in a way that allows for the development and maintenance of social cohesion. The conventionalist can note that this psychological need for stable mechanisms ensures that we will seek (implicitly or explicitly) conventions of a particular sort; this need ensures that *any* set of conventions that fulfills the function of morality will contain certain broad families of norms. Though the discussion that follows will be brief, it will be sufficient to demonstrate that the conventionalist has the makings for a plausible response to the worries about the scope of his account.¹⁶

First, in order for one to have personal projects (as opposed to mere wishes or fantasies), one must have access to certain resources, and have at least some reasonable degree of comfort—in the sense of being able to make reliable

¹⁶ Wong (2006: chaps. 2, 7) provides an elegant and more elaborate version of the examples that follow. The argument here also finds support in Sugden 1986/2001. Furthermore, recent work in economics and psychology (e.g., Frank 1988, Haidt 2003) demonstrates how human emotions and affective responses function to stabilize and solidify our conventions.

predictions—that one will be able to use and develop those resources over time. Since norms about the use, development, and distribution of resources would best provide this, they will be a part of any set of conventions that fulfills the function of morality. For similar reasons, there will be norms prohibiting certain harms. It's common knowledge that personal projects can conflict. In light of this, to be able to pursue personal projects (as opposed to merely having or planning them), one needs comfort that, when there is conflict, one's projects will not be unnecessarily or arbitrarily limited. Having norms prohibiting unnecessary or arbitrary harms would do this. So, again, we can expect any set of moral conventions to include harm norms. For similar reasons, it's plausible that there will be norms requiring restitution if one's projects are unjustly spoiled. The need for norms regarding harms and restitution for injustices seem particularly important if one is to be willing to engage in longer-term projects.

We can also expect that there will be norms concerning reciprocity and assistance. Successful cooperative activities—especially those like Hume's grain harvesting example in which one party must commit to providing his services before the other party does—are not possible without a general tendency toward reciprocity. But this tendency is too easily frustrated in the face of individual self-interest. Because it's so easy for self-interest to spoil cooperative ventures, we can expect moral conventions to include norms requiring reciprocity. For largely similar reasons, we can expect that there will be norms requiring assistance or beneficence for those in need. As we have seen, we must often rely on others—individuals or institutions—to help us achieve our personal projects. But, left unchecked, the pursuit of individual self-

interest can prevent there from being the degree of social cohesion and trust necessary for this support network to be possible. So any set of moral conventions will also include some sort of norms requiring assistance to those in need. Though this account would need to be further fleshed out, I take it to be sufficient to demonstrate the conventionalist has the resources to give a plausible account of the scope of morality.¹⁷

(1.3) Moral Conventionalism, Subjectivism, & Relativism. We're now in a position to say something about the differences among moral conventionalism, subjectivism, and relativism. We've seen that conventionalism identifies moral facts with certain social conventions. This means that moral facts are fixed at the level of *groups* or *cultures*.¹⁸ Subjectivism, by contrast, takes moral facts to be fixed by the subjective judgments of particular *individuals*. Thus, conventionalism and subjectivism are competing meta-ethical accounts. But they are both compatible (and typically combined with) relativism. To see how this is so, notice that relativism contrasts, not with conventionalism or subjectivism, but rather moral absolutism—the view that there is a single true morality that holds for all groups and individuals. Though it's possible that all groups and individuals could come to the same constitutive conclusions about the moral facts, most subjectivists and relativists (including Harman and Wong) deny this. That is, most subjectivists and conventionalists are also relativists.

¹⁷ In future work I hope to further explore the ability of a conventionalist account of this sort (e.g., Wong 2006) to provide a robust account of the demands of beneficence and justice.

¹⁸ See, for instance, Wong 2006: 71, Harman 1977: chap. 9, Harman & Thomson 1996: 22-4.

2. The Allure of Moral Conventionalism

Now that we have a basic sense of what moral conventionalism is, we can see why it makes for an appealing meta-ethic. For one, by taking moral facts to be a function of established conventions, the conventionalist gets the makings of an intuitively plausible account of moral error: a moral judgment is incorrect just in case it is inconsistent with the relevant moral conventions. He also gets an account of moral facts that is naturalistically acceptable and metaphysically conservative: a convention, after all, is the upshot of nothing more than a certain shared conditional preference. Furthermore, this shared conditional preference—namely, that each individual in a group G chooses to adopt a particular solution S given that the others in G do so as well—provides the conventionalist with a naturalistic account of the normativity of morality. At a gloss, he can explain the ways in which a moral norm (convention) is reason-giving: It's because a given norm N satisfies my conditional preference that I have a reason to follow the solution it specifies. And it's because others share this conditional preference with me that gives them a reason to expect and want me to follow N , and so a reason to hold me accountable if I fail to meet their expectations and desires.

Moreover, by drawing on the above functional account of conventions (§1.2), the conventionalist can give a principled explanation for why we take moral considerations to have greater authority than those of (say) etiquette. Moral considerations are considerations that concern the adjudication of competing personal projects in a manner that preserves social cohesion. Considerations of etiquette, by

contrast, concern the facilitating of social interaction. Thus, it is because moral conventions serve a more psychologically significant function (e.g., preserving personal projects versus facilitating social interaction) that they have greater authority.¹⁹

A further benefit of a conventionalist account of morality is that it seems particularly well positioned to explain the diversity of moral rules that exist between—even within—different societies and cultures. Because conventions (moral or otherwise) arise as solutions to the distinctive coordination problems of particular groups, we can expect that different groups will endorse different conventions. But to the extent that different groups face similar types of coordination problems, we can also expect that they will adopt similar solutions to those problems. Recall the case of money: While different societies have adopted different currencies—dollars, yen, euros—the benefits of having a convenient, easy to carry medium of exchange helps explain why, despite these differences, different societies have all come to adopt some form of paper currency.

The discussion of the last two sections demonstrates that there is much to like in moral conventionalism. Not only does it appear well positioned to make sense of the fallibility, normativity, and content that we take morality to have, it does so in a way that's both naturalistically acceptable and metaphysically conservative. In fact, there is much in the sophisticated conventionalism of Harman and, especially, Wong

¹⁹ Clearly, more would need to be said to explain the conventionalists' account of normativity. But this will be sufficient for present purposes.

that I believe is largely right—the account of our moral psychology and the functional demarcation of the moral, in particular. However, though I believe that moral conventions play an essential role in a viable account of moral facts, I will argue below that they cannot be the whole story.

3. Conventionalism and Moral Error

Despite the allure of conventionalism, further reflection on what conventions are, and how they get established provides us with good reason to think that the conventionalist cannot provide a plausible account of moral error. To make this argument, I'll first demonstrate that because our actual moral conventions are often incomplete, inconsistent, and based on false beliefs, they are unsuitable as the basis for an account of moral error. I'll then show that any conventionalist move to address these problems by appealing to some “corrected” set of conventions requires him to give up what is distinctive and interesting about his meta-ethical position.

(3.1) The Need to Focus on “Corrected” Conventions. In this sub-section, I develop three examples that demonstrate that our actual conventions prove incapable—at least on their own—of providing a plausible account of moral error. For each of these examples, it would be possible for the conventionalist to accept the problematic results. However, doing so for any one—or, worse, for all—would come at a significant cost. A better option, it would seem, lies in taking moral facts to be fixed, not by our *actual* conventions, but rather some suitably *corrected* set of them.

First, as is readily apparent, we often face new and novel moral problems—for instance, we've recently been confronted by questions about the permissibility of

performing medical research on cloned human stem-cells. However, if moral facts are fixed by our actual conventions, there cannot be a moral fact of the matter about whether such research is permissible. As we've just seen, conventions only exist given that a group faces a *recurring* coordination problem. But stem-cell research is a *new* problem. So there cannot be—by definition—a convention, and so cannot be a moral fact. While the conventionalist might be able to provide an account for how to extend existing conventions to cover new cases (I'll explore the plausibility of that option shortly), the point here is that existing conventions cannot—as they are—be held to fix the facts for new moral issues. This result is problematic for the conventionalist's understanding of moral discourse: unless more is said, not only is moral error *impossible* in such cases, but any anxiety one has about making the right moral choice, or guilt one might feel for deciding as one does, is *irrational*—there is, after all, nothing one could be wrong about.

A further deficiency arises because conventions, as we've seen, are fixed at the group level, and so the facts applicable to individual moral judgments are determined by the conventions of the relevant group. But, as is often the case, a given individual will be a member of different groups, and those groups will endorse conflicting conventions for particular issues. In such cases, it will be (metaphysically) indeterminate as to which convention fixes the facts. Consider an example: Is it morally permissible for Jane, a Democrat who is also a Catholic, to support her party's pro-choice congressional candidate? Democrats and Catholics have clear answers to this question. So the conventionalist can say that it's permissible for Jane *qua*

Democrat, and impermissible for her *qua* Catholic. But suppose Jane takes herself to be a Catholic-Democrat (rather than a Democrat who happens to be a Catholic, or a Catholic who happens to be a Democrat). The conventionalist can say nothing about the (im)permissibility of Jane's vote *qua* Catholic-Democrat. Since there's no convention, there cannot be a moral fact. So we could never say that Jane's judgment was incorrect. And, as with the first case, we must take any anxiety or guilt that Jane feels about her choice to be irrational. Thus we have another broad set of situations where conventions prove incapable of providing us with moral facts, and so incapable of explaining how moral error is possible.²⁰

Here, the conventionalist might be willing to just accept the results—Jane faces a tragic, but alas familiar, situation where she must choose which group is more important to her. However, this move is problematic for a variety of reasons.²¹ First, it takes what, intuitively, we regard as a moral issues (e.g., is it morally permissible for me to support a pro-abortion candidate?) and makes it a practical one (e.g., which group do I identify with more?). Second, biting the bullet here is more plausible to the extent that cases involving conflicting conventions are relatively uncommon. But given the panoply of group affiliations that most persons have, there's no reason to think conflicts like Jane's are unusual. The result is that the conventionalist is

²⁰ Note that a parallel worry arises when there is conflict between the conventions endorsed *within* a particular group. Suppose that I've been wronged. Suppose further that in my group there are both conventions requiring that justice be served, and conventions demanding forgiveness and mercy. What, on a conventionalist account, is the fact of the matter regarding whether I should forgive?

²¹ Similar considerations apply to a conventionalist move to bite the bullet with regard to the problem of new moral issues just discussed.

committed to morality being much less determinate than our commonsense thinking suggests it is.

Finally, there is nothing in the basic conventionalist account to prevent conventions from being based on false beliefs. Moreover, such a result is not a fanciful possibility, after all, antebellum American conventions that permitted slavery were based (in part) on false beliefs about their genetic inferiority of blacks. But if moral facts are fixed by actual conventions, then not only is the conventionalist committed to repugnant conclusions about the permissibility of slavery in antebellum America, he is also committed to there being no necessary connection between a moral fact and the truth of the (non-moral) beliefs that undergird it. The cost of accepting this is significant—it amounts to the skeptical claim that morality is nothing but a convenient projection.

(3.2) The Problem of Specifying “Corrected” Conventions. Taken together, these examples demonstrate that our *actual* conventions cannot form the basis for a plausible account of moral error. So it should not be surprising to see that most sophisticated conventionalists do not take moral facts to be determined by them, but rather by some set of “*corrected*” conventions.²² More specifically, conventionalists have tended to adopt two general strategies—hypothetical and non-hypothetical—for explaining what this “corrected” set of conventions is. But as we will see, both commit the conventionalist to an unpleasant dilemma: they can specify a plausible set of

²² See, for instance, Harman 1977: 94-5, Harman & Thomson 1996: 14, Wong 1984: chap. 4; 2006: 71.

corrected conventions only if they give up the distinctive and interesting features of their meta-ethical account.

(3.2.1) Non-hypothetical Strategies. Given their commitments to naturalism and metaphysical conservatism, conventionalists often seek to articulate what the relevant corrected conventions are in a way that does not appeal to hypothetical situations or tools of idealization. Here the dominant proposal holds that there are higher-order conventions that specify what to do in situations where one faces a new problem or a set of conflicting conventions.²³ But this proposal doesn't resolve the problem; it merely pushes it off. When we consider what these higher-order conventions might be, we find not only that there are too many candidates, but that their prescriptions are incompatible. To see this, return to the case of Jane. On the going proposal, there should be a higher-order convention of (roughly) the form: when confronted with incompatible conventions, do *D*. The problem is that there are too many different and conflicting ways to fill out the content of *D*. Should Jane flip a coin, consult an expert or confidant, go with her gut feeling, do something else? Is there supposed to be an even higher-order convention that selects between these options?²⁴ The conventionalist needs to provide answers, but given how few constraints there are in his general framework, what could ground his proposal?²⁵

²³ See Harman 1977: 95 and Wong 1984: 32, 39.

²⁴ Notice that a seemingly plausible candidate—namely, decide on the balance of reasons—is too vague to help.

²⁵ I assume the given the conventionalist's naturalism and metaphysical austerity, he would be unwilling for make a move to a robust (e.g., Kantian) account of rationality, or a faculty of intuition.

(3.2.2) Hypothetical Strategies. The problems with the non-hypothetical strategy suggest that conventionalists would fair better if they explained “corrected” conventions as the upshot of some sort of hypothetical or idealized procedure. The most common proposal is one that takes the relevant conventions to be those that would result from a suitably informed and rational group of individuals.²⁶ The suggestion is that under such idealized conditions we would get sufficiently determinate conventions—the parties would become aware of their false beliefs, know how to make sense of new cases, and know how to decide between conflicting conventions.

But if this proposal is to work, the conventionalist would need to address the significant worries that are generally thought to undermine appeals to this sort of idealization: What does it mean to be “suitably informed”? Is there an answer to this question that doesn’t require an implausibly radical form of idealization?²⁷ Moreover, can these questions be answered without positing some set of robust hypothetical conditions (e.g., full information, a veil of ignorance) that would conflict with the conventionalist’s underlying commitment to metaphysical conservatism? Even if there were a coherent account of what it is to be “suitably informed,” why suppose that the

²⁶ Harman (1977: 95-6, 131, Harman & Thomson 1996: 14) seems to make a suggestion of this sort, as does Wong (1984: 39-40).

²⁷ See Veleman 1988 and Sobel 1994. It’s worth noting a tension in Harman’s conventionalism on this point. While he at times seems inclined toward idealization as a way of addressing the determinacy concern (e.g., 1977: 95-6, 131, 1996: 14), he also raises doubts similar to those in the text about its plausibility (e.g., 1977: 111-2).

decisions of some idealized set of individuals could have normative force for us here and now?²⁸

A better option—one that would allow the conventionalist to avoid (or address) these questions—would seem to be one that appeals to some hypothetical, *but non-idealized*, agreement procedure as the way to correct our actual conventions. But here too there are problems. First, the appeal to some hypothetical scenario would seem to trivialize the role of *conventions* in the conventionalist’s account; all the substantive work would be done by the hypothetical conditions. Because of this, it is hard to see why the resulting account should be considered a variety of conventionalism, and not some variety of hypothetical agreement constructivism (chap. 4.3). Second, the normativity worry that plagues idealized accounts applies here too: why suppose that the decisions reached under some hypothetical, though non-idealized, circumstances could have normative force? After all, the hypothetical scenario isn’t our scenario, so what reason is there to think we’d accept its pronouncements as binding?²⁹

Finally, in the absence of any idealization, there is no reason to think that the move to a hypothetical agreement addresses the original three problems. It does nothing to ensure that the resulting corrected conventions (the purported “moral facts”) are based on true non-moral beliefs. But we also have no reason to think that—without idealization—we would come to any agreement about how to either deal with

²⁸ See Rosati 1995 for an argument of this sort.

²⁹ See the discussion (chap. 4.3.2.2) regarding the problems with Gauthier’s (1985) hypothetical agreement account.

new moral problems, of resolve conflicting conventions.³⁰ Unless the conventionalist can come up with some other option, hypothetical strategies will also prove unable to provide the needed account of “corrected” conventions.³¹

(3.3) Why this is a Problem Conventionalists Cannot Accept. The above discussion reveals that the move to see moral facts as fixed by our actual conventions leads to an implausible account of moral error. But it also reveals that the conventionalist cannot, at least not without incurring unacceptable costs, address the deficiencies of our current conventions by proposing that moral facts are fixed by some “corrected” set of conventions—there’s just no plausible story about how one might move from our flawed, actual conventions to a suitably corrected set. This gives us a powerful reason to reject moral conventionalism. But it’s worth noting that the *conventionalist himself* is committed to this conclusion. As we have seen, the conventionalist takes the function of morality to be that of providing us with stable structures for the assessment and adjudication of conflicting personal projects. This means that he is committed to a skeptical account of morality if it turns out that there

³⁰ This may be too strong. In the next chapter, I develop a counterfactual test that a conventionalist might be able to use to secure the right sort of tie between moral conventions and the non-moral beliefs on which they are based. However, even were this possible, there’s no reason to think it would help address the other problems discussed in the text.

³¹ A final possibility: The conventionalist might seek to address the limitations of our actual conventions by endorsing moral particularism. This move would allow the conventionalist a potentially more plausible way to accept the claim that our actual moral conventions are pervasively indeterminate (the makings for this sort of move can be found in Harman 1977, 2005; but see Wong 2006: 23-35). However, such a move comes with significant challenges. For one, the conventionalist must provide a plausible argument for the particularism—no small feat. Furthermore, standard particularist accounts take one of two forms: (i) moral facts are identified with *non-natural* facts (e.g., Dancy 1993); (ii) moral facts are taken to be a function of the judgments of *competent moral agents* (Wiggins 1987, McDowell 1985). Either option poses a hurdle as the conventionalist is forced to demonstrate that the appeal to particularism doesn’t collapse into non-naturalism, subjectivism, or some sort of ideal observer account.

are no such structures. Though the conventionalist aims to provide a *non-skeptical* account, the above problems demonstrate that he cannot.

This point merits emphasis. We've just seen that the best the conventionalist can provide is a set of moral conventions (and so a set of moral "facts") that is in various ways incomplete, inconsistent, and based on false beliefs. But surely such a set of conventions is incapable of providing anything close to the sort the structures that are necessary to fulfill the function of morality: Because so much of our substantive moral inquiry concerns how to address new problems and how to make sense of conflicting moral norms, the fact that our moral conventions can give us little—if any—direction on these issues demonstrates that conventions alone are insufficient to fulfill the function of morality.³² So, on his own terms, the conventionalist is unable to avoid a slide into skepticism.³³

Despite the above arguments, a conventionalist might still try to maintain that our moral conventions, imperfect though they are, still work well enough.³⁴ Given the extent of the problems we've uncovered, there's good reason to question the plausibility of this line of reply. But, setting that aside, how might the conventionalist substantiate this proposal? The best option would be to appeal to language (or

³² It's worth noting that Wong criticizes absolutist moral theories because they "contribute very little to the resolution of genuinely controversial normative issues. They give either vague advice that fails to favor one side or another or advice that is reasonably rejected outright [by the other side]" (2006: 25-6). If—as seems reasonable—this is the relevant standard for fulfilling the function of morality, then it's hard to see how a conventionalist like Wong could meet it.

³³ Notice that the conventionalist cannot give up the functional claim without undermining his argument that he can both secure a plausible account of the content of morality, and distinguish moral conventions from non-moral ones (§1).

³⁴ In correspondence, both David Wong and Jamie Dreier suggested this as a possible line of response.

etiquette) in the following way.³⁵ Linguistic conventions are flawed in the same ways that moral conventions are: our talk of sunrises is based on false beliefs; our use of words and phrases can be indeterminate ('bank', 'He's cool'); and our conventions can be inconsistent—e.g., conventions requiring that we avoid ambiguity and obscurity can conflict with conventions to be brief.³⁶ But these features of our linguistic conventions don't keep them from doing their job. We still manage to communicate effectively because, when our linguistic conventions run out, we have alternative, pragmatic ways to keep communication going. In light of these facts about language and communication, it would be a mistake to think morality is any different.³⁷

But morality *is* different. To see why, we need to take a closer look at the case of language. A distinctive feature of our linguistic practice is that we are generally willing to accept pragmatic solutions in order to make up for the limitations of our linguistic conventions: we accept the other's locution even if we don't like it; we stipulate a more precise word or phrase to overcome an ambiguity. Now if morality functions just like language as the conventionalist's proposal maintains, then this predicts that—when our moral conventions run out—we should be willing to accept some pragmatic compromise. But that's exactly *not* what we find. As we have seen, when our moral conventions break down, something significant is typically at stake

³⁵ Let's grant for the sake of argument that conventionalist accounts of language (and etiquette) are correct.

³⁶ Grice 1989: 27).

³⁷ Both Harman (1977: 113) and Wong (2006: chaps. 2, 9) see the limits of our moral conventions as something that is to be fixed in pragmatic manner (i.e., in a way that relies on something other than our moral conventions).

(personal projects, social cohesion). Because of this, we are generally much less willing to accept pragmatic solutions. When we take something of moral significance to be at issue, we tend to think that compromise is inappropriate; we're inclined to stand our ground and are even willing to carve others out of our moral community if necessary. Consider: Republicans and Democrats who have become independents have often done so *because* they've found their former party's pragmatic solutions to moral issues unsettling. This is significant. It reveals that the conventionalist's analogy fails. The very features that allow imperfect linguistic conventions to, nonetheless, fulfill their function of facilitating effective communication are missing in the case of morality. If we're inclined to stand our ground or carve others out of our moral circles when our conventions breakdown, then conventions have proven unable to fulfill their function; they have proven unable to provided us with stable mechanism for the assessment and adjudication of competing personal projects in a way that preserves social cohesion.³⁸

4. The Subjectivist Turn

The above arguments indicate that conventionalism alone does not have the resources to provide a plausible account of moral facts and moral error. The thought

³⁸ Here is a good a place as any to note a further implausible implication of the conventionalist's proposal—one concerning moral inquiry. On the conventionalist's account, when we're contemplating the permissibility of a new moral issue like stem-cell research we are—strictly speaking—not engaged in inquiry about what the moral facts are (as a new moral issue, there just isn't one). Rather, we're doing something else. For instance, according to Harman and Wong, we're arguing about what the new convention should be (Harman & Thomson 1996: 29-31; Wong 2006: ch. 9). So, while intuitively we take moral inquiry to be a process of *discovery*, it is—by the conventionalist's account—really a process of *creation*.

that this deficiency could be addressed by an appeal to some set of “corrected” conventions proves either to be unhelpful, or to clash with the underlying motivations of conventionalism. But this does not mean that there is nothing of value in the conventionalist’s proposal. The conventionalist’s insight lies in locating the core of morality in a distinctive psychological need—namely, the need for a stable set of mechanisms capable of protecting the pursuit of personal projects and social cohesion. It’s this need after all that drives us (implicitly or explicitly) to develop and follow moral conventions. But he over-estimates what conventions are capable of doing. In particular, he is wrong to think that they are sufficient as the basis for an account of our moral discourse. But if conventions don’t fix the moral facts, what does?

Enter subjectivism. My proposal starts with the conventionalist’s picture of our psychology. But I do not take moral facts to be constituted by the resulting conventions. Rather, I take moral facts to be facts about the judgments that would bring an individual to a distinctive point of psychological stability. At a gloss, the stability in question is one in which the individual in question has come to a point of equilibrium between (i) the psychological significance that she places in her personal projects, and (ii) the psychological importance that she sees in the maintenance of social cohesion. As we will see, the move to subjectivism allows us to bring in additional resources that not only fix the problems in the conventionalist’s account of moral error, but also preserve what’s distinctive and interesting in his account—namely, a naturalistically acceptable and metaphysically conservative account of the normativity of morality.

Chapter 6: Sophisticated Subjectivism

Neither Harman nor I have discussed subjectivism, since hardly anybody nowadays thinks it plausible. Judith Thomson (1996: 199)

The discussion of the previous chapters suggests that moral facts are subjectively validated moral conventions. This chapter aims to develop and defend this idea. My particular proposal holds that moral facts are facts about the set of corrected moral conventions (corrected in the sense of being precisified, expanded, disambiguated, and tied to the non-moral facts in the right way) that one would stand by as valid given a motivation to secure stable structures for the assessment and adjudication of competing personal projects. As will become apparent, this proposal delivers an account of the objectivity of ethics that is procedural, response-dependent, and metaphysically austere.

The form of my argument will be as follows: After some work to locate my proposal in logical space (§1), I turn to explain the essential—but limited—role of moral conventions. Given what we’ve learned about their limitations (chap. 5), I argue that they should be granted only defeasible status as moral facts (§2). In order to identify the set of conventions that actually fixes the moral facts, I turn to investigate the nature and purpose of moral judgment: getting an understanding of the social-psychological mechanisms that drive moral judgment allows us to identify a subjective procedure the execution of which specifies the constitutive set of moral conventions (§§3-5). I then conclude by demonstrating that my proposal delivers a plausible account of the normativity and fallibility that are distinctive of morality (§§6-8).

1. Orientations

As mentioned above, my account takes moral facts to be constituted by a metaphysically austere, response-dependent procedure. But since these distinctions cut across the taxonomy of the earlier chapters, let me briefly explain them. Procedural accounts (e.g., Kantianism, ideal observer accounts, hypothetical agreement views) take moral facts to be fixed through the execution of some procedure or the passing of some test. By contrast, non-procedural accounts (e.g., realism, simple subjectivist views) take facts to be the upshot of something else: for realists it's conforming to some mind-independent feature of the world, for the simple subjectivist it's the making of a particular judgment. Among procedural accounts, there are response-dependent and response-independent varieties. Response-dependent versions hold that subjective responses (e.g., what one would judge to merit approval, or what one would want after cognitive psycho-therapy) play an ineliminable and constitutive role in the procedure; on response-*independent* versions (e.g., Kantianism, many varieties of hypothetical agreement accounts), subjective responses do not play an essential role. Finally, metaphysically robust versions of response-dependent proceduralism employ strong forms of idealization/abstraction or employ a substantive notion of agency/personhood (e.g., Brandt's requirement that one become *fully informed* during cognitive psycho-therapy, Korsgaard's appeal to a *Kantian conception* of persons). Austere versions, by contrast, do not make use of such assumptions.

In light of these distinctions, we get the following rough characterization of my proposal:

A convention M is a moral fact (for S) just in case M would provoke response R in S under conditions C ,

where (i) C articulates a procedure or set of tests, (ii) R specifies a distinctive response or syndrome of behavior, and (iii) neither (i) nor (ii) makes use of metaphysically robust materials. As this characterization makes plain, my proposal entails that moral facts are relative to the responses of agents. Because of this, we can see my proposal as a sophisticated variety of moral subjectivism. With our orientation now complete, let's turn to the details.

2. The Place of Moral Conventions

As we have seen (chap. 5.3), it is implausible to *identify* moral facts with our actual moral conventions—it is too often the case that they are incomplete, inconsistent, or based on false beliefs.¹ That said, it is reasonable nonetheless to take them as having *defeasible* status as moral facts.² For one, we've seen that given reasonable, empirically well-founded claims about the nature of human psychology, the conventionalist framework delivers an intuitively plausible (though vague) account

¹ Some reminders on terminology: (1) Throughout this chapter I will use 'moral fact' to refer to the truth-makers of our moral judgments; my usage is intended to be neutral on what the underlying metaphysics is (e.g., the phrase alone does not imply anything about whether or not moral facts mind-independent). (2) I will use 'moral convention' in the manner of chap. 5: moral conventions are self-enforcing regularities of behavior that serve as solutions to moral coordination problems. I do not, however, endorse the (conventionalists') further claim that moral facts just are these conventions—the story, as we'll see, is more complicated than that. (3) I will use 'moral judgment' to refer not merely to the act of asserting things like 'I ought to give to OxFam,' but to the larger syndrome of behavior of which such assertions are a part.

² It's worth noting that this claim (or something very much like it) is endorsed by many "stark raving moral realists" (Railton's term). Though want an account that delivers a robust notion of fallibility, they also maintain that our moral beliefs and conventions must be *approximately true*. See, for instance, Sturgeon 1985, Boyd 1988, Brink 1989: chap. 5, and Railton 1986a, 1989.

of the content of morality: prohibitions against harms, property rights, and (limited) demands for assistance and reciprocity (chap. 5.1.2). Furthermore, given that morality is essentially concerned with providing us with stable mechanisms for the assessment and adjudication of competing personal projects, we should grant anything that fulfills this function at least *prima facie* status as a moral fact. But since moral conventions are often reasonable (though vague) satisfiers of this function, they pass the test. So even though our moral conventions can be insufficiently determinate, inconsistent, and based on false beliefs, they still function to specify a set of general moral principles that not only form the basis for moral deliberation and inquiry, but have defeasible status as moral facts.³

The challenge is to identify a way of moving from an imperfect set of conventions, to a set that provides a plausible account of moral facts. In particular, we're looking for a way of correcting our existing moral conventions that will allow us to capture the distinctive features of our moral discourse (e.g., fallibility, normativity), but that does not rely on robust metaphysical claims. Because moral judgment provides a tangible and familiar entry point to the nature of morality, and in light of the conclusions we reached at the end of chapter 4, I propose to investigate the question of what moral facts are by looking at the nature of, and psychological mechanisms involved in, moral judgment.

³ The points in this paragraph do not undermine the conclusion from chap. 5 that our moral conventions alone fail to provide a viable account of moral error. Rather, we can see our moral conventions as providing something akin to the rough but clearly incomplete and imperfect account that folk psychology provides us about our mental life.

3. The Nature and Purpose of Moral Judgment

Moral judgment is a distinctive form of assessment. As we have seen, it is an assessment of the appropriateness of a particular action as a response to a given moral problem (where moral problems are understood as the upshot of competing personal projects). But because making a moral judgment requires one to take a stand on a particular solution to a problem, making one will tend to provoke a distinctive and familiar form of anxiety—what I will call m-anxiety.⁴ M-anxiety is a cognitive and affectively prompted unease about the legitimacy of one's assessment that a particular action is an appropriate response to (solution for) a given moral problem. Because m-anxiety is provoked by the clash of personal projects that is distinctive of moral problems, it can arise either in the course of making a moral judgment, or when one reconsiders—on reflection or when pressed—a moral judgments that one has already made.

To flesh this out, consider first-personal moral judgments—judgments that an individual makes about what she ought to do. To make such a judgment is to take a stand on matters concerning the assessment or adjudication of conflicting personal projects. Because such a judgment takes a stand on the relative merits of competing personal projects, it has the potential to bring psychological distress. After all, one faces a situation where one's own personal projects conflict with those of others. So

⁴ As will become apparent, the unfamiliarity of the label 'm-anxiety' does not undermine the claim that the psychological state to which it refers is familiar. Rather, it just reflects the fact that English lacks a term for the state in question (though something like 'conscience' may come close).

something must give: either one must concede one's personal projects (and so accept the psychological distress that will come with this), or one must stand one's ground on the respective merits of one's project (and thereby accept that disruption to social cohesion that this will bring). Either option tends to come with a psychological cost, a cost that takes the form of m-anxiety.⁵

We can say more about what m-anxiety is and what triggers it. First, though m-anxiety will tend to be more common in cases where one faces a moral situation that is unfamiliar or non-routine, it can be elicited even in familiar and routine cases so long as the issue at hand is presented in a novel or forceful manner. Second, what triggers m-anxiety, and how intensely it manifests itself, will vary across individuals. Work by Jonathan Haidt and his colleagues on incest supports both of these points.⁶ Typically, judging that it is wrong for siblings to make love with one another does not prompt m-anxiety—it's a judgment whose truth one tends to be quite confident about. However, Haidt's work demonstrates that one's judgments about the wrongness of such acts will prompt m-anxiety if one is made to consider a case where the obvious reasons against incest (e.g., that it can lead to birth defects) have been ruled out (e.g., two forms of birth control were used). Moreover, Haidt also notes that subjects felt varying levels of

⁵ I believe that there is a similar psychological dimension to third-person moral judgment—judgments that an individual makes about what *others* ought to do. To the extent that one has an actual connection to the individuals that one is making a judgment about, or—more generally—to the extent that one can put oneself in their shoes, one will experience a tension between the significance that one sees in the pursuit of personal projects, and the significance that one sees in the preservation of social cohesion. For the sake of simplicity, I will focus just on first-person moral judgments in what follows.

⁶ See, for instance, Haidt 2001 and Haidt, et. al. 1993.

anxiety when they were prompted to defend their judgments about incest in these cases.

While the above picture is intuitive and familiar, it prompts two high-level empirical conjectures: First, it suggests that m-anxiety is a (nearly) universal feature of the psychology of social beings like us.⁷ This is meant as the modest claim that we have an evolved capacity to experience m-anxiety; it is *not* the implausibly strong claim that m-anxiety manifests itself on every occasion of moral judgment.⁸ We've also seen that m-anxiety is a cognitive and affectively prompted unease about the legitimacy of one's assessment of the appropriateness of a given action as a response to a particular moral issue. But since it involves conflict between *both* cognitive and affective elements, it is not happily understood as either a type of belief, or an emotional response. Rather, and this is the second conjecture, it's best understood as a high-order mental state that combines elements of each.⁹

⁷ Sociopaths—as individuals incapable of feeling m-anxiety—are not, in the relevant sense, social beings. Hence, our practice (clinically and non-clinically) of referring to them as ‘antisocial’. There is likely also a social disorder at the other extreme—individuals that are so hyper-sensitive to clashing personal projects that they are unable to function socially. If there are such persons, they too would seem to fail to be, in the relevant sense, social beings.

⁸ Two further points: (1) That we have an evolved capacity to experience m-anxiety receives some empirical support from recent on the anxiety baboons experience in the face of (perceived) disturbances to their social hierarchies. See, Cheney & Seyfarth 2008. (2) My conjecture is a *descriptive* claim—namely, that social beings like us will be disposed to experience m-anxiety; it is not a *normative* claim about whether, or to what extent, we *should* feel m-anxiety.

⁹ Two points: (1) There are other mental states that appear to have this structure. Consider two examples: (i) *schadenfreud*—a state that combines *feelings* of delight with *judgments* that another is suffering; (ii) general social anxieties which often have both a cognitive dimension (e.g., a belief that others don't like you) and an affective one (e.g., general feelings of inadequacy or embarrassment). (2) Viewing m-anxiety as a higher-order mental state is consistent with granting a place for more familiar (lower-order) moral emotions like anger, guilt, and resentment in moral judgments and m-anxiety. However, I do not think that the standard moral emotions can be the whole story; we need m-anxiety as well. For one, as the example of first-person moral judgment indicates, m-anxiety will not always

If something like the above account of m-anxiety is right, it has significant implications. For starters, recognizing that m-anxiety is an evolved response in social beings like us prompts a couple of questions. First, we can ask what (adaptive) function this response serves. That is, what role does m-anxiety play in the lives of social beings such that it would be advantageous for them to expend resources feeling it and seeking to alleviate it? Second, given an account of the function that m-anxiety serves, we can ask how this function ought to be fulfilled. That is, what ways of alleviating m-anxiety would comport with the role that it has evolved to play? An answer to these questions will—I believe—provide us with an account of the normativity (in a teleological sense) of moral judgment that points to a distinctive account of moral facts.¹⁰

Turning to the first question, I maintain that m-anxiety functions as a uniquely effective tool for the maintenance and promotion of social cohesion: It functions to check the self-interested pursuit of personal projects by sensitizing individuals (in a manner to be explained) to the disruptive effects that the pursuit of their personal projects could have for various social relationships and institutions. The underlying

manifest itself as one of these distinctive emotions, but rather a more general angst. Second, we can experience m-anxiety by merely considering our moral beliefs (see, for instance, Aronson's (1999) empirical work on certain forms of cognitive dissonance). I hope to give a more complete account of m-anxiety, including these two points, in future work.

¹⁰ Given my critical arguments concerning sentimentalist accounts that appeal to basic emotions (chap. 4.1), one might think that my broadly evolutionary account of m-anxiety can't be right. Aren't I positing a basic emotional/affective state that psychologists don't recognize? Doesn't that make my account of moral facts *ad hoc*? Though I feel the force of the this concern, I believe it is mistaken. For one, the state that I'm describing is not a *basic* emotion, but rather a higher-order state that is constituted (in part) by basic emotions (see point (2) of the previous note). Second, as will become apparent, unlike the sentimentalists, I do not take m-anxiety to be constitutive of the moral facts. So there is nothing *ad hoc* in what I'm doing.

thought here is intuitive and familiar, especially given the discussion of moral psychology in chapter 5: Social relations and institutions make it possible for individuals to pursue projects that they wouldn't otherwise be able to (e.g., cultivating a skill, raising a child, launching a business venture). Because these personal projects give meaning and purpose to our lives, we have strong, self-interested incentives to pursue them. In fact, these incentives can persist even when pursuing them compromises the ability of others to pursue their own personal projects. This can lead to problems: left unchecked, the self-interested pursuit of personal projects by individuals can erode the very social structures that make those projects possible. The upshot is that the preservation and maintenance of social cohesion requires one to at times forego one's personal projects so that others may pursue their own.

It's here that m-anxiety plays its distinctive role: It serves as a regulating device—it makes one aware (in a sense to be explained) of the costs that one's pursuit of one's own self-interested personal projects will have on others in a way that dampens one's self-interested inclinations. Extending work by Robert Frank (1988) suggests how this dampening might work.¹¹ Frank argues that certain emotions (e.g., anger, guilt, love) function to *change* one's incentive structure away from purely self-interested pursuits. Feelings of guilt are unpleasant. Because guilt is provoked when one considers acts like lying and stealing, it tends to make those acts unappealing even in situations where one knows that one won't get caught doing them. This happens

¹¹ A complementary proposal can be found in Haidt 2001.

because guilt affects a change in one's incentive structure; it makes acts that could be in one's self-interest (stealing when one knows one won't get caught) unappealing.

Though my account of the function of m-anxiety does not require me to take a stand on how the underlying psychological mechanisms work, I'm inclined to think that something analogous to Frank's proposal must be on the right track: Just like guilt functions to change our incentive structure so that stealing becomes unappealing, m-anxiety changes our orientation to our personal projects. It sensitizes us to the costs that our (self-interested) pursuit of our own personal projects will have on others, and does so in a way that dampens our (self-interested) inclinations. M-anxiety—in a sense—changes the significance that our personal projects have for us in a way that facilitates the maintenance and promotion of social cohesion.¹²

If m-anxiety is a form of psychological distress that is provoked by concerns about whether one's moral judgment inappropriately privileges one's own self-interest, and if we have developed a sensitivity to it as a tool for facilitating the maintenance and promotion of social cohesion, we can ask (and this is the second question from above) how it should be addressed. That is, what means for the alleviation of m-anxiety would comport with the (adaptive) function that it plays? The account of moral judgment and moral psychology we've developed suggests that m-anxiety is properly alleviated by demonstrating (to oneself and others) that one's moral judgments do not give one's own interests undue preference—that they are, in a

¹² I develop this point further within the context of promises in Kurth ms.3. In future work, I hope to examine the differences between my affectively based proposal and Gauthier's (1985) rationalist alternative.

sense, *valid*.¹³ Here's why: M-anxiety is an unease about the legitimacy of one's moral judgments that we've developed in order to foster social cohesion (and so secure its benefits). It sensitizes us to the concerns of others. To escape this unease, we need comfort that our moral judgments appropriately accommodate the concerns of others. Getting this comfort—in a manner that comports with the function of m-anxiety—requires validating our moral judgments.¹⁴

We now have an account of the nature and purpose of moral judgment—one explained in terms of m-anxiety and validation. We get some confirmation that this account is right in virtue of its explanatory power. For starters, we've seen that m-anxiety functions as a mechanism for social sensitization. Noting that it serves this function helps explain how self-interested beings like us could develop and sustain robust social networks. Moreover, given the benefits that come with the ability to maintain and promote social cohesion, we have an explanation for why we have the ability to experience m-anxiety. Finally, the account also helps explain our disposition to seek to validate our moral judgments—it helps explain why we're so concerned about the justifications of the moral judgments that we make.

¹³ As will become apparent, I am not using the logician's notion of validity here.

¹⁴ It's worth noting that there is an ambiguity in 'alleviate'. On one understanding, it refers to processes that *address the source* of the anxiety; on another reading, it refers to actions that just make m-anxiety (appear to) go away. I'm using 'alleviate' in the first sense. Noting this explains why drugs, hypnosis, denial, and the like are not really ways of alleviating (in the first sense) m-anxiety. Rather, they are ways of making it (appear to) go away by burying, masking, or blocking it.

4. Validation: Its Nature and Function

M-anxiety sensitizes us to the concerns of others, and it motivates us to validate our moral judgments—to demonstrate that we are not giving our own interests undue preference. But what must validation be like in order for it to be capable of alleviating m-anxiety in the right way—that is, in a way that comports with the function that it serves (call this a *fitting* validation)? In what follows, I'll argue that fitting validation has two dimensions—one concerning truth and another involving empathy. But I will also show that there are certain structural challenges that frustrate our ability to demonstrate that these two dimensions of validation have been met. This will put us in a position to see the appeal and power of a response-dependent procedural account of moral facts.

(4.1) Dimensions of Validation: Truth and Empathy. If m-anxiety is prompted by concerns about the appropriateness of one's moral judgments, then these concerns are in part driven by questions about whether one's moral judgment is based on an accurate understanding of the non-moral features of the problem one faces. Moreover, if m-anxiety is a tool for the maintenance and promotion of social cohesion, then this function is more likely to be fulfilled to the extent that one can demonstrate that one's judgment is not based on false non-moral beliefs. So one's validation will be fitting to the extent that one can demonstrate to oneself and others that the non-moral claims underlying one's moral judgments are true and that the inferences one draws from them are sound—this is the truth dimension of fitting validation. While the claim that fitting validations will be tied to considerations of truth has the air of a platitude, it

will be worthwhile to consider some illustrative examples: We regard another's moral judgments as invalid, as mere dogmatism or confusion, to the extent that they are unable to ground them in the non-moral features of the situation at hand; we reject the racist's moral judgments about minorities at least in part because we take them to be based on false (non-moral) claims about the genetic inferiority of the groups he derides.

We can see the second principal driver of m-anxiety as a family of related concerns—concerns that, for instance, in making a moral judgment one has not properly identified the parties who are affected by the situation at hand, or that one has not taken the situations of these affected parties in to proper consideration. Again, this has the air of a platitude—complaints that you have failed to consider the situation of some group, or that claim that you have failed to understand the plight of an affected party, will tend to provoke m-anxiety in you. So your validation will be fitting to the extent that you can demonstrate to yourself and others that you have identified all the affected parties, and that you have properly taken their situations into consideration. Because such demonstrations involve—in a sense—putting yourself in the shoes of others, we can refer to it as the empathy dimension of fitting validation.¹⁵

(4.2) Structural Challenges. Though demonstrating that one's moral judgment is not based on false non-moral beliefs, and that it properly takes the situations of all affected parties into consideration can be difficult, the challenge is particularly

¹⁵ I take questions about whether a moral judgment is based on faulty moral reasoning to be captured by the empathy dimension of fitting validation.

difficult with regard to the empathy dimension. There are two reasons for this. The first is an upshot of the limitations inherent in our moral conventions. As we've seen, viable demonstrations of empathy involve, among other things, showing that one has properly identified who the affected parties are. But who one identifies as an affected party turns, in part, on the criterion one uses for determining what counts as (say) a harm or an injustice. Similarly, how one weighs the harms to one party against the harms to another will be a function of what one takes the relevant standard to be: Must one consider the situation of all parties impartially? If so, should one use an aggregative conception or a non-aggregative one? Now here's the problem: though it's true that for a given moral issue there will be a network of conventions that help us identify candidate criteria for identifying affected parties and assessing their situations, these considerations will, as we've seen (chap 5.3), too often provide content that is either insufficiently determinate or inconsistent. But without determinate criteria, how can there be a fact of the matter about what a viable demonstration of empathy involves?

Second, there is a fundamental asymmetry between one's assessment of the significance of one's own personal projects, and one's assessment of the significance of the personal projects of others that introduces another structural barrier to demonstrations of empathy. This is because as we develop our personal projects, we become psychologically invested in them in the sense that their role in our mental lives grows in significance; but since it's not possible for others to be similarly invested, they cannot assess the significance of those projects in a similar manner. We

can make this vivid with an example: My personal project to complete my dissertation has the significance that it does because of the psychological and intellectual investments that *I* have made—because of the role that it plays in *my* mental life. If you have a similar personal project, or if you try to imagine what it would be like to be in my shoes, you might be able to get *some* sense of the significance of my project. But you cannot see my project *as I do*—you cannot see it as *my* project. At best, you see my project as *you imagine that I do*.¹⁶

The problem here is not merely epistemic—it's not just the result of it being difficult to understand first-person assessments from a third-person perspective. Rather, the point is that many of the things that make a personal project significant are essentially bound up in the perspective of the individual whose project it is. These assessments of significance have an ineliminable subjective element to them.¹⁷ Clearly, this fundamental asymmetry frustrates successful demonstrations of empathy: How can I demonstrate to you that I've taken your situation as an affected party into sufficient consideration if it's not possible for me to fully appreciate your situation?

¹⁶ While this asymmetry claim is not uncontroversial, it is endorsed by a broad range of moral philosophers—e.g., Nagel (1979), Dancy (1993: 188), Gauthier (1985: chap. 3), Gibbard (1990), Hampton (1993), and Harman (1977). In future work, I hope to say more about this asymmetry, and the value of personal projects more generally.

¹⁷ It's worth noting that the argument here does not depend on an implausible claim that first-person awareness is *infallible*. Rather, it merely requires that individuals have *some* authority with regard to the nature of their mental lives. Though this more modest claim is somewhat controversial (see, for instance, Boghossian 1989), it is the received view among philosophers of mind (e.g., Moran 2001, Burge 1996). I defend the legitimacy of modest first-person authority in Kurth 2003.

Though you might agree that my assessment is “close enough,” you might not. And if you don’t, there seems to be no way for us to come to a *principled* resolution.¹⁸

(4.3) Subjective Validation. These two structural challenges are real impediments to our ability to demonstrate that we have properly addressed the empathy dimension. However, we’re often times able to come to acceptable resolutions to moral problems *despite* these challenges. Moreover, I think we can gain insights as to the nature of the procedure that is constitutive of the moral facts by looking at how this happens—how it is that we, at least sometimes, succeed in demonstrating that our moral judgments are valid despite these challenges.

In actual practice, fitting validation is a balancing act in the following sense: In seeking to validate a moral judgment, one seeks to demonstrate to oneself and others that the truth and empathy dimensions have been addressed. To do this, one appeals to both particular features of the situation in question, and the associated network of conventions in an effort to show that one’s moral judgment is not based on false non-moral beliefs, that it correctly identifies the affected parties, and that it properly considers their situations. One takes one’s moral judgment to be valid to the extent that one feels that there are features of the situation that allow one to make a compelling case that the truth and empathy dimensions have been satisfied. But

¹⁸ A couple of points: (1) Notice that an appeal to some impartial point of view does not help provide a principled resolution to our disagreement—an impartial point of view is, after all, nothing more than a fancy type of third-person perspective. So there’s no reason to think that it can make first-person assessment transparent. (2) Notice as well that this asymmetry helps explain the common thought that normative issues are essentially contestable. This is significant given the extent to which claims of essential contestability are used in meta-ethical debates (e.g., Mackie 1977, Wiggins 1987, D’Arms 2006).

because validation involves a psychological balancing of this sort, and because this balancing takes place within the mental economy of the judging agent, fitting validation will—in practice—have a substantive, subjective element to it.

But that it is subjective does not entail that it is arbitrary. Here's why: M-anxiety, as we've learned, sensitizes one to the concerns of the affected parties. So consider what happens if one's reasons are bad—if they inaccurately represent the non-moral features of the situation, or if they overlook some feature of an affected party's plight. One will be unlikely to mollify the affected parties' concerns, and so do little to alleviate the m-anxiety that one is experiencing. By contrast, good reasons will tend to have some ability to assuage the concerns of others, and so will tend to bring one at least some relief. This is doubly significant: It means that we can measure the success of one's demonstration of the validity of one's judgment in terms of the ability of one's reasons to reduce the m-anxiety that one feels. And it demonstrates that a subjective validation procedure needn't be arbitrary.

Before moving on, let's pause to review the principle conclusions that we have reached so far. First, we have identified a response—m-anxiety—that appears to play a central role in raising our awareness to, and helping us solve, the types of coordination problems that are distinctive of morality. Second, we have developed a broadly teleological account of how to properly alleviate m-anxiety—namely, a validation procedure with truth and empathy dimensions. Third, we have noted that there is good reason to think that fitting validation is possible despite the structural problems that we identified with regard to true and empathy. Finally, we have noted

that while the truth dimension of fitting validation is likely to have “objective” satisfaction conditions, the satisfaction conditions for the empathy dimension will likely involve a psychological balancing that is (somewhat) subjective. These conclusions are significant because, as we will see, they point us to a novel but powerful account of moral facts.

5. A Procedure for Fixing the Moral Facts

Stepping back, we saw in §2 that our moral conventions are legitimate candidates for moral facts. We also noted that they would actually constitute the facts only if they could be corrected. Now we have just learned that in actual practice we take our moral judgments to be valid to the extent that we feel there is a compelling case that the truth and empathy dimensions have been adequately addressed. I think that these insights, when combined with what we have already learned about moral judgment and moral psychology, reveal a distinctive, procedural account of moral facts.

(5.1) The Proposal. The discussion of the previous section suggests that the correct set of conventions—the set that is constitutive of the moral facts—could be the upshot of a procedure that embraces truth and empathy dimensions of fitting validations, and a subjective balancing mechanism that I will call “standing by”. The proposal I would like to explore takes moral facts to be moral conventions that (i) are not the result of false non-moral beliefs, and that (ii) one would stand by as properly assessing the situations of all affected parties. This needs fleshing out, so let’s consider

condition (i)—the truth requirement—and condition (ii)—the empathy requirement—more closely.

(5.1.1) The Truth Requirement. As both commonsense and the above discussion of the truth dimension of fitting validations draws out, there is a necessary connection between the moral facts, and the truth of the non-moral beliefs on which they are based. However, providing a non-trivial account of this connection proves to be a challenge. Notice, for instance, that it would be implausible to just rule out any moral convention that is based on false non-moral beliefs. After all, there are surely moral conventions that are intuitively legitimate even though they are—as a matter of contingent fact—the result of false beliefs. For instance, we don't want to rule out conventions obligating us to help others just because they were, one way or another, the result of superstition or false non-moral beliefs about the nature of universe. But we also want to be able to capture the intuition that part of what makes conventions sanctioning discrimination wrong is that they are based on false beliefs about (say) the genetic inferiority of the persecuted group.¹⁹

Our general understanding of conventions suggests a way to explain the tie between the validity of a moral convention and the truth of the non-moral beliefs on which it is based. In light of the account of conventions we've been using, whether a particular solution *S* to a given coordination problem becomes a convention is

¹⁹ It's worth noting that moral conventionalists like Harman (1975, 1977) and Wong (1984, 2006) overlook this challenge (or fail to appreciate its significance): though they talk of the need to make our actual conventions responsive to the non-moral facts on which they are based, they say almost nothing about how this is supposed to be done.

determined by how salient S is in comparison with the alternatives—the more salient, the more likely it is to become a convention.²⁰ Suppose S becomes a convention at least in part because the false belief b made S more salient. We can ask whether S would have become a convention were the false belief b corrected. If it would not have, then we know b played a substantive role in establishing S as a convention. For this reason, it makes sense to think of S as illegitimate.

This account of how conventions get established suggests that the tie between the validity of a moral convention and the truth of the non-moral beliefs on which it is based should be understood counterfactually. More specifically, to assess whether an existing moral convention is based on false non-moral beliefs in a manner that compromises its legitimacy, we ask whether that convention could have been established were we to correct the false beliefs that it (actually) resulted from. If the convention would have persisted, then the false beliefs do not undermine the legitimacy of the convention.

There are a few points to emphasize about this counterfactual test. First, the test only lets in conventions whose connection to true beliefs was robust enough to make it more salient than the alternatives. So it gives us a procedure that ensures there will be a necessary connection between moral facts and the truth of the non-moral beliefs on which they rest. Furthermore, we see that assessing these counterfactuals does not require us to consider *why* the false belief b made option S more salient, or

²⁰ See, for instance, Lewis 1969 and Sugden 1986/2001.

whether it *should have*. We just need to know that it *did*. But all things considered, that's a fairly straightforward empirical question.²¹

(5.1.2) The Empathy Requirement. In our actual deliberations, we attempt to satisfy the empathy dimension by seeking a distinctive sort of psychological equilibrium—we try to minimize the m-anxiety that we're experiencing by demonstrating that our moral judgments do not give undue preference to our own interests (§4.3). I think something like this response—the securing of a distinctive psychological equilibrium I've called standing by—*just is* what it is to meet the empathy requirement. But we can be more specific about the essential features of the standing by response. It involves both a distinctive pattern of behavior and a distinctive motivation. First, one is *disposed* to defend the grounds of one's judgment on reflection and when pressed; one will appeal to features of the situation in question (e.g., associated moral conventions, one's own moral rules of thumb and paradigm examples, non-moral facts) in an effort to demonstrate that one has correctly identified the affected parties and properly assessed their situations. Second, one approaches one's defense *motivated* to secure stable structures for the assessment and adjudication of competing personal projects. So a moral convention meets the empathy requirement to the extent that one would stand by it—i.e., to the extent that one would defend it (on

²¹ Given the argument of chap. 5, it's worth emphasizing that even though a moral conventionalist could make use of my counterfactual test to rule out conventions based on false beliefs, this is still not enough to salvage the view. As we have seen, even factually accurate moral conventions will be too incomplete and too (potentially) inconstant to be plausible candidates for the moral facts. In fact, it's in part because of this that we also need the empathy requirement.

reflection and when pressed) given that one is motivated to secure stable structures for the assessment and adjudication of competing personal projects.

For reasons analogous to those noted above (§4.3), meeting the empathy requirement is not arbitrary or random. Rather it is shaped and constrained by one's capacity for m-anxiety. M-anxiety, as we have seen, functions as a regulating device—it makes one aware of the costs that one's pursuit of one's personal projects will have on others in a way that dampens one's self-interested inclinations. So securing the psychological equilibrium that is distinctive of the state of standing by will involve the judging agent taking the concerns of the affected parties seriously. This in turn will tend to result in her changing her orientation toward her own personal projects so as to make room for others to pursue theirs.

To flesh this out, let's return to an example from chapter 5—the case of Jane, the Catholic-Democrat, and her question about whether it is morally permissible for her to support her party's pro-choice congressional candidate. To simplify matters, let's assume that the relevant conventions meet the truth requirement. What then does meeting the empathy requirement involve? For starters, Jane must demonstrate to herself and others that in coming to her conclusion (say, that supporting the candidate is permissible) she has correctly identified the affected parties, and that she has properly assessed their situations. Among other things, this will involve identifying a set of criteria for specifying who the affected parties are (e.g., a harm principle) and a method for assessing their situations (e.g., a non-aggregative notion of impartiality) that is responsive to the concerns of those who will be affected by her decision. Call

this the *grounds* for her judgment. Moreover, it must also be the case that Jane would—given the needed motivation—stand by her judgment. That is, it must be the case that she would affirm the grounds of her judgment on reflection and when pressed by others. Success here constitutes meeting of the empathy requirement.

But why think that *standing by* a moral judgment is constitutive of what it is to *actually meet* the empathy requirement? A significant part of the answer is that the pattern of behavior and motivation that standing by involves fits with the account of the nature and function of m-anxiety that we have developed: M-anxiety sensitizes us to the concerns of others so that we may address conflicting personal projects in ways that tend to preserve social cohesion. That's its adaptive function. Moreover, m-anxiety is alleviated (in a way that corresponds to this function) only when one exhibits the pattern of behavior and motivation that is constitutive of standing by one's assessment that the empathy requirement has been met.²²

Bringing the parts of this discussion together gives us the following account of moral facts:

Moral facts: A convention M is a moral fact (for S) just in case M (i) meets the counterfactual test of the truth requirement, and (ii) is something S would stand by—i.e., something S would defend (on reflection and when pressed) given that S is motivated to secure stable structures for the assessment and adjudication of competing personal projects.

²² I believe that there are interesting parallels between the proposal in the text and Allan Gibbard's (1990: chaps. 1-5) influential account of normative governance. But I also think that my proposal has distinct advantages—e.g., it is not tied to an expressivist account of moral judgment, it points to something more robust than conversational norms as the psychological driver of our moral inquiry and discussion. I hope to explore these idea further in future work.

(5.2) Objections and Replies. But despite the tight fit between the function of m-anxiety and my account of how m-anxiety is to be properly alleviated, one might object to my account of moral facts. I now turn to take up three of the more pressing concerns one might have (others will be addressed in the sections that follow).

(5.2.1) Metaphysical Austerity Lost. The first objection starts by noting that my account takes moral facts to be fixed in virtue of a *hypothetical* procedure—a procedure that allows us to abstract away from, among other things, the structural challenges inherent in the empathy requirement, and that relies on a robust counterfactual test. It then charges that my appeal to these procedures not only compromises the aim of providing a metaphysically austere account of moral facts, but opens my account up to criticism of the sort I leveled at idealization and conventionalist accounts.

The general concern here could be specified in a variety of ways. Given the critical discussion of the earlier chapters, I will take up what I see as the two most significant versions. First, one might maintain that the procedure I employ for correcting false beliefs (or any of the other abstractions I use) involves an implausibly radical form of idealization—one that goes beyond what is (psychologically) possible.²³ But as noted in §5.1.1, the counterfactual test I propose for correcting false beliefs is modest and empirically tractable. Individuals do not need to be made fully and vividly aware of all the facts; they don't need to consider them from a variety of

²³ A concern of this general form can be found in Velleman 1988 and Sobel 1994.

points of view and across a spectrum of moods; nor is there any demand for perfect reasoning. Rather, my proposal employs only the social-psychological mechanisms that we *actually* use to come to conventional solutions to (moral) coordination problems.

But this answer might seem to run me head on into a second form of the objection: My procedures for meeting the truth and empathy requirements privilege—in an *ad hoc* manner—a specific point of view. In particular, they privilege a certain set of subjective responses.²⁴ Though I acknowledge that my account privileges certain subjective responses, I deny that it does so in an unprincipled manner. For starters, one of the primary lessons of chapters 3-4 was that the robustly objectivist views considered there failed *because* they sought to establish a foundation for morality that went beyond individual subjective responses. Moreover, much of this chapter has been devoted to demonstrating that though my account of moral facts might seem revisionary, it is also rich in explanatory power and intuitive plausibility.²⁵

(5.2.2) Circularity. A second objection charges that my proposal is viciously circular:²⁶ My account of moral facts makes appeal to notions of *moral* conventions, non-*moral* beliefs, *affected* parties, and *proper* assessment. But because the account relies on this set of inter-related moral notions, it is both uninformative and

²⁴ An objection of this sort is advanced in Sobel 1994.

²⁵ There are two further ways of fleshing out this concern. The first focuses on matters of normativity: why should one think that my *hypothetical* procedure would have any (motivational or reason giving) force for us here and now? The second concerns the content of morality: why think that my hypothetical procedures will be capable of delivering an intuitively plausible account of the content of morality? These are legitimate worries. I take them up in detail in §§6-7 below.

²⁶ We've seen (chap. 4.3) that this is a problem that can undermine constructivist accounts.

inaccessible to anyone without an antecedent understanding of these key terms. However, while it's true that my account explains moral facts in terms of other moral notions, this is not a problem for the account also provides a way to get into the circle. To see this, notice that my account explains the function of morality, not by appeal to other moral notions, but rather in terms of what's needed for the adjudication of competing personal projects. Moreover, I've already argued (chap. 5.1) that these phenomena—personal projects, their tendency to conflict, and the need for a stable way to resolve these conflicts—are familiar features of our lives. So we have an independent way to access the moral terms that my account uses in virtue of the personal projects each of us has and pursues. And the fact that my account is independently accessible means that it is also informative. So the circularity charge does not stick.²⁷

(5.2.3) Genuine Disagreement. A final objection: Because my account takes moral facts to be fixed by individual judgments, one might maintain it will be unable to explain how any two people can share a common notion of (say) 'wrong'. If so, it will be unable to explain how genuine moral disagreement is possible. The key to recognizing the problem with this concern lies in recognizing the *essential* role that existing moral conventions play in structuring and framing our moral discussions. As we've seen, moral argument and moral disagreement *must* take place against this

²⁷ One might try to revive the circularity worry by arguing that this functional account is too vague to allow us to draw a sharp line between the moral and the non-moral. Suppose this is true; it is neither surprising nor problematic. After all, moral language—like virtually all language—is vague. Moreover, this vagueness is not debilitating since there is strong, cross-cultural agreement on paradigmatic instances of the moral and the non-moral (see Haidt 2006).

backdrop. So interlocutors will have conceptual schemes that ensure that they are not (generally) talking past one another.²⁸

(5.3) The Verdict. With these objections laid to rest, we have the makings of a metaphysically austere, response-dependent, procedural account of moral facts. Granted, this account of moral facts represents a departure from the standard thinking about what grounds the objectivity of ethics. But in addition to learning (in chaps. 2-5) that this standard thinking is deeply flawed, we have (in this chapter) identified a variety of features that count in favor of my proposal. First, it draws on an empirically informed account of human psychology. Moreover, as we've seen, this psychological base provides a unified explanation of moral judgment, moral psychology, and the point and purpose of morality. Second, by identifying moral facts with conventions that are not based on false beliefs, it secures a tolerably revisionist account of the content of morality (more on this below). Finally, because it is built from a metaphysically austere base, it avoids the troubles that undermined accounts (realist and constructivist) that relied on more robust assumptions, and does so within a descriptivist account of our moral discourse.

While I take these features to provide us with a compelling, *prima facie* case in favor of the view, there is still work to be done. I must still show that this form of

²⁸ A couple of points: (1) It would be unreasonable to demand perfect overlap between two persons' concepts of 'wrong'. Moreover, the pragmatic account of, for instance, Lewis 1979 demonstrates how substantive argument and disagreement are possible even given a significant degree of difference in conceptual schemes. (2) The proposal in the text is not the only way that a subjectivist could make sense of disagreement. He could also follow conventionalists like Wong (1984, 2006) and Harman (Harman & Thomson 1996) in appealing to an expressive function of moral judgment as what unites disagreeing parties.

sophisticated subjectivism provides a plausible overall account of our moral discourse. In particular, I need to demonstrate that my subjectivist proposal can accommodate the normativity and fallibility we associate with moral judgment. To the extent that these tasks are met, we will have additional reasons to endorse my proposal.

6. Fallibility

My account of moral facts provides a robust account of moral error. For starters, as the above discussion of Jane's predicament indicates, one's actual moral judgments, and the judgments that would meet the truth and empathy requirements, can and do come apart. More importantly, the discussion to this point makes plain that the moral judgments one actually makes are the upshot of a rich, complicated, but inchoate set of features. They are the product of things like (i) what one takes the relevant moral conventions to be, (ii) one's education and life experiences, and (iii) one's cognitive, affective, and conative endowment. In light of this, it should not be surprising that one will not stand by all of one's (actual) moral judgments. Being challenged by another, or reflecting on one's judgment, can lead one to back off. It might, for instance, provide one with new reasons, or reveal that portions of the grounds for one's assessment are in tension with one another.

But because my proposal takes moral facts to be the corrected conventions that an individual agent would stand by, one might object that the resulting account of error will be implausibly relativistic in the sense that it must allow for the possibility that intuitively appalling conventions might count as morally valid. This objection comes in two forms. The first version holds that, on my account, an otherwise good

person could be a member of a society that has endorsed monstrous conventions. But if he, for no fault of his own, does not realize that things could be morally otherwise, he would likely stand by the monstrous conventions of his society. This means that, for instance, the sheltered racist's judgments would be moral facts (for him). The second form of this objection charges that my account must accept that the judgments of moral monsters are (for them) morally right. After all, such individuals could be psychologically constituted such that they would stand by monstrous conventions.²⁹ But having to accept either result would mean that my proposal requires intolerable revisions to our commonsense thinking about morality.³⁰

I think that the above discussion gives us very good reason to conclude that appalling moral conventions would not pass the truth requirement (§5.1.1). It is reasonable to think such conventions are invariably rooted in false non-moral beliefs.³¹

²⁹ If the monster is a full-fledged sociopath (i.e., someone incapable of feeling m-anxiety), I think there are compelling grounds for ruling him out as a moral agent (see note 7 above). But presumably non-sociopathic-but-still-monstrous-individuals are possible. I take the objection in the text to be raising a worry about this type of person.

³⁰ There is a further, related objection one might make. One might argue that my account of moral facts is inherently unstable in the following way: My account holds that moral fact about a given issue is the upshot of (among other things) the judgment that an agent would stand by—the judgment that secures a distinctive psychological stability. But the judgment that brings this stability for one issue, might lead to instability for another. So, for instance, when Jane comes to a point psychological stability with regard to the permissibility of abortion, this might have a destabilizing effect on her psychology with regard to (say) capital punishment. But because there is nothing in my proposal to rule out such possibilities, my account of moral facts threatens to be inherently unstable. As will become apparent, the argument I give to address the sheltered racist and moral monster worries also allows me to address this worry.

³¹ This claim is supported by recent research on how our disgust response can be co-opted in a manner that leads to the establishment of intuitively appalling moral conventions. When an individual has the false belief that the members of some group are in some way “dirty” or “contaminated,” research shows that she will tend to associate feelings of disgust with those individuals. But this affective tie leads her to develop a disposition to take members of the “dirty” groups to actually *be* disgusting. The result is that practices that persecute these groups become more salient. Once this happens, it becomes much easier for conventions sanctioning the persecution of these groups to take hold. The counterfactual test, in combination with this empirical work, gives us a principled way of ruling out moral conventions that

But for the sake of argument, let's grant—generously—that both the sheltered racist's and the moral monster's judgments pass the truth requirement. To see how my subjectivist proposal is still able to address these objections, we need to start with an understanding of why my account seems vulnerable to them in the first place. To do this, notice that the satisfaction conditions for the empathy requirement are not sensitive to the judging agent's ability to conceive of alternative solutions to the situation at hand, or to her ability to imagine how the various choices she makes will affect others; that is, the satisfaction conditions are not sensitive to what I will call the “moral ingenuity” of the judging agent.³² The problem is that agents display varying degrees of moral ingenuity: To the extent that an agent cannot see that alternative courses of action are possible, or is limited in her ability to understand a situation from alternative perspectives, the judgments she stands by will tend to conflict with what we take to be the intuitively right answers. But because my proposal doesn't (or couldn't) give an account of what counts as a sufficient amount of moral ingenuity, it is saddled with the significant cost of taking the judgments of the sheltered racist and the moral monster as constitutive of the moral facts.

It's undeniable that we have strong intuitions that the judgments of the racist and the monster are just plain wrong. This means my proposal is less plausible to the

have intuitively appalling content—e.g., racist, persecutory, or otherwise discriminatory conventions—in virtue of the role that the underlying false beliefs played in establishing them. See, for instance, Kelly ms., Lewis 1969, Nichols 2004, D'Arms & Jacobson 2006, and Haidt 2006.

³² Moral ingenuity, as I am understanding it, has both a cognitive dimension (e.g., one's ability to *reason* about thought experiments) and an affective dimension (e.g., one's capacity for feeling what it's like to be in another's shoes).

extent that it can't capture these intuitions. But I can give an account of our intuitions about the racist and the monster that explains why we react so strongly to them. The first thing to notice is that my critical arguments of chapters 3-4 entail that we cannot be assessing the racist and the monster with some "robustly objective" standard of moral ingenuity—morality doesn't involve such things. So what are we doing? We're assessing the moral ingenuity involved in the judgments of the racist and the monster as *comparatively less objective* with regard to some alternative (e.g., our own level, the level of our society, that of some moral exemplar). But because there is no "objective" or "universally acceptable" level of moral ingenuity, we cannot specify a unique standard of measurement. Instead, we should take moral facts to be relative to some contextually fixed level of moral ingenuity.³³

Here's the upshot: while we must allow that there is a sense in which the moral judgments of sheltered racist and the moral monster are true (for them), we also have a clear and intuitively powerful explanation for why their moral judgments are deficient—they employ an impoverished level of moral ingenuity. But we can say more. We can explain why the racist and the monster employ degrees of moral ingenuity that are *bad*, *deficient*, or otherwise *impoverished* ones, and not ones that are merely *different* than ours. To do this, we need to have a better understanding of what moral ingenuity is. Moral ingenuity is a capacity that people can have in the same way that they can have capacities for being political savvy or a capacity to be creative.

³³ The nature of this context sensitivity is a complicated matter. Though I will not address it here, I hope to explore it in future work.

What these capacities have in common is that they only admit of characterizations that are loose and open-ended. These characterizations are loose in the sense that we can only explain them with platitudes: To be politically savvy is to be, among other things, cunning, astute, and crafty. To be creative is to be imaginative, innovative, and original. To have moral imagination is to be sensitive to the concerns of others, resourceful in one's comparisons, and ingenious in identifying alternatives. These characterizations are open-ended in that their content tends to shift with both time and circumstances.

This is significant in three ways. First, it explains why these capacities can be exemplified in such strikingly different ways. Lyndon Johnson and Mohandas Gandhi were both paradigmatically politically savvy; both Albert Einstein and James Joyce were stunningly creative; Socrates and Jesus evinced tremendous moral ingenuity. Second, as these examples draw out, that we only have loose characterizations of these capacities does not mean that we're wrong to think that these exemplars of political savvy, creativity, and moral ingenuity were better than—and not merely different from—their peers. This is because our characterizations, loose and open-ended though they may be, still have structure that is broadly teleological. Consider: we understand that Johnson and Gandhi were more politically savvy than their peers when we see how much better they were in their ability to identify and manipulate sources of power in order to advance the protection of civil rights. Similarly, one is more morally ingenious to the extent that one can identify structures that will provide particularly stable mechanisms for the assessment and adjudication of competing personal

projects. One's moral ingenuity will be a function of, among other things, structural features such as one's ability to get a sense for what it is like to be in another's shoes, and one's facility for working through thought experiments.

Finally, though these structural features mark out how one person's use of a capacity is *better* than another's, they do not provide a way of specifying a particular manifestation as *the best*—the characteristic features of these capacities are just too vague and open-ended for that. What's the optimal set of skills to have to be a savvy politician? To be creative? To be morally ingenuous? Given what we've learned, we can see that such questions are just ill-formed. Because of this, we can also see that the thought that my account cannot explain why (say) the racist's capacity for moral ingenuity is deficient is misplaced. So my account delivers a robust notion of moral fallibility.

7. Normativity

One might worry that given the inchoate and fluid nature of the beliefs and attitudes that make up the point of view from which we make and assess moral judgments, my subjectivist proposal will be incapable of explaining the distinctive normativity that we take moral considerations to have. Assessing this concern requires that we have an understanding of what's distinctive of the normativity of morality. Though it's debatable whether the following four theses exhaust our intuitive sense of the normativity of morality, I believe they capture what many take to be its core:

Universality: If morality demands that you Φ in C , then *anyone* in a situation like C must also Φ .

Categorical: The demands of morality apply *regardless* of one's endorsement of them.

Authority: Moral considerations *trump* non-moral considerations (e.g., considerations of self-interest, etiquette, aesthetics)

Stability: Moral considerations provide a *fixed* normative guide in the sense that they are a central and stable regulative force in our lives.

I have stated these theses in their strongest form. Though some maintain that the normativity of morality entails these very strong claims, others—myself included—do not. In the present context, rather than argue against these overly strong formulations, it will be more helpful to demonstrate the significant extent to which my proposal can approach them.³⁴

(7.1) Universality. The accounts of moral psychology and moral conventions that I endorse work to ensure that *any* valid moral system—that is, any set of moral conventions that meet the truth and empathy requirements—will include certain general moral prohibitions, permissions, and demands. More specifically, all valid moral systems will include prohibitions against harms, provisions for property rights, and (limited) demands for assistance and reciprocity (chap. 5.1). In this way, my proposal accommodates a central intuition underlying the universality thesis—namely, that we are *all* subject to the same moral requirements. But because my account allows that the specifics of these general requirements can differ, it does not capture the strong version of the universality thesis. I do not see this as a cost, but rather a

³⁴ The general and quick version of my argument against these strong formulations is that they involve an illegitimate inference from the observation of some *robust pattern* in our moral discourse, to the conclusion that there is some *essential feature* underlying that pattern that is constitutive of the normativity of morality.

benefit—it allows us to explain our competing intuitions about whether moral demands are universal.³⁵

(7.2) Categorical. My account secures the categorical thesis in the sense that it agrees that moral considerations apply independently of our *actual endorsement* of them. But they are not completely endorsement independent—moral demands must be something that one *would endorse* (stand by) as having met the empathy requirement given that one is motivated to secure stable structures for the adjudication of competing personal projects. Again, I see this moderation as a point in favor of my account. It accommodates the competing intuitions of morality as something whose normative force is external to us, and as something that could not be wholly divorced from our individual aims and aspirations.

(7.3) Authority & Stability. On my account moral considerations get their distinctive authority and stability in virtue of the fact that moral considerations have a psychological role that is deep and wide.³⁶ Here's what this means: a mental state—a belief, aim, desire, feeling, etc.—has a *deep* psychological role just in case it would be extremely costly, if even possible, to rid oneself of that mental state. For instance, feelings of joy and anger seem to have deep psychological roles given how extremely difficult it would be to cause oneself to never feel them. A mental state has a *wide* psychological role just in case it is prevalent across broad swaths of one's mental life.

³⁵ Against the strong rendering of the universality thesis, it is worth noting that recent empirical work indicates that the folk are not universalists—they allow that different groups/individuals may be subject to different moral requirements. See, for instance, Goodwin & Darley 2008.

³⁶ C.f., D'Arms & Jacobson 2006.

Again, the role of anger and joy also seems to be wide since these feelings can be had with regard to just about any aspect of one's life. In this way, both our attachment to our personal projects and our aversion to m-anxiety are like anger and joy. They have a psychological role that is both deep and wide. It's this feature that explains how they're able to provide us with especially *stable* normative guidance.³⁷ And it's this feature that explains why we generally treat moral considerations as having special significance (in comparison with considerations of, say, self-interest, etiquette, or aesthetics). In general, moral considerations have special significance for us because our pursuit of our personal projects, and our desire for social cohesion, shape and inform so much of how we think and what we do.³⁸

In sum, though my account does not capture the strongest form of the core dimensions of the normativity of morality, it gets close. Moreover, as I have tried to suggest, what I can't capture is unlikely to extract much of a cost given the inchoate state of the underlying intuitions.

³⁷ One could augment this explanation in the text by noting that many of our moral norms (e.g., prohibitions against harms, demands for equality) have the central role that they do because they are intimately tied to our affective and emotional response systems. See Frank 1988, Haidt 2001, 2006, and Nichols 2004 for details.

³⁸ Noting the psychological significance of our attraction to our personal projects and our aversion to m-anxiety helps address concerns from Rosati (1995) and others. The worry is that views like mine that ground the normativity of morality in some *hypothetical* endorsement fail to be able to explain why more considerations have force us *here and now*. But in light of the discussion in the text, we see that the normativity of morality is—on my account—grounded in more than mere hypothetical endorsement. That it is also grounded in aspects of our *actual* psychology—namely, the psychological significance of our personal projects—explains why the Rosati-style objection has little force.

8. A Better Account of Moral Facts

In lieu of a conclusion, I close by comparing my account to the moral conventionalism considered in the last chapter. First, though my account carves out an essential role for conventions in structuring and framing moral decision and inquiry, the moral facts are not constituted by a procedure located at the level of groups or cultures as they are on the conventionalists' account. Rather, moral facts are fixed at the level of individuals. This feature is significant for it allows my account to avoid two problems that undermine conventionalism—namely, an intolerably indeterminate account of moral facts, and an implausible account of moral inquiry. Let's take these in turn.

First, because a conventionalist account of moral facts can appeal to only group/culture level phenomena, it is required to conclude—implausibly—that there are no moral facts in cases that either present us with new moral issues, or that involve conflicting conventions (chap. 5.3). By contrast, my account brings in the individual, and so has access to a greater set of resources for dealing with these sorts of cases (e.g., life experiences, rules of thumb). But one might worry that my account, nonetheless, still proves to be insufficiently determinate. I think this is mistaken, but showing why will take some work.

We can begin by noting that, on my account, there will be no moral fact of the matter only in cases where either the truth or empathy requirement isn't met. So in order to assess whether my proposal proves to be insufficiently determinate, we need to understand what sorts of situations would keep either condition from being met.

First, consider the truth requirement. It ensures that the moral conventions that are constitutive of the moral facts are tied to truth in the right way. This means that in order for the truth requirement to fail in a way that would entail that there was no moral fact for a given issue, *all* the moral conventions applicable to that issue would need to fail the counterfactual test discussed above (§5.1.1). While this scenario can't be ruled out, it also represents such an extremely skeptical view that I think we must set it aside. Moreover, were such a skeptical scenario actual, it would cut equally against my proposal and any conventionalist alternative.

Let's turn to consider what would have to happen to prevent the empathy requirement from being met. Since meeting the empathy requirement involves standing by a particular moral judgment, failing to meet the empathy requirement would require a situation in which there was *no* moral judgment that one would defend while being motivated to secure stable structures for the assessment and adjudication of competing personal projects. True moral dilemmas—predicaments so awful that one is unable to make a choice of what to do—seem like the only way this could happen. Moreover, we've seen that conventionalists face a similar problem. So it's hard to see how having to acknowledge possible failures of this sort makes my account problematically indeterminate.

A more troubling possibility is that there are *too many* ways for an agent to meet the empathy requirement—that is, one might be willing to stand by multiple judgments for a given moral issue. That is, for a given moral question, one would both stand by a judgment that says one has an obligation to Φ , and stand by a judgment that

says one has an obligation to Ψ , where Φ -ing and Ψ -ing are not compossible. Thus my account allows a form of indeterminacy that threatens to bring inconsistency or even incoherence.

But the force of this concern diminishes once we again take note of the role of moral ingenuity. First, the fact that (for a given moral issue) one might stand by judgments that are not compossible is *prima facie* evidence that the something is deficient in one's moral reasoning or moral understanding.³⁹ It indicates a need for greater moral ingenuity. But once we recognize this, we can see that greater moral ingenuity is likely to resolve the conflicting judgments about what one morally ought to do. One rather obvious possibility is that we come to see that we don't have obligations to both Φ *and* Ψ , but rather an obligation to Φ *or* Ψ .⁴⁰ The upshot is that though my account does entail that some moral questions may be indeterminate, it—unlike the conventionalist account—does so in a manner that better comports with our intuitions.

With regard to moral inquiry, recall (chap. 5.3) that on the conventionalist's account, when we're contemplating the permissibility of a new moral issue like stem-cell research we are—strictly speaking—not engaged in inquiry about what the moral facts are (as a new moral issue, there just isn't one). Rather, we're doing something else. For instance, according to Harman and Wong, we're arguing about what the new

³⁹ Sidgwick (1877: 6) makes a similar point.

⁴⁰ Sayre-McCord (ms.) makes a similar proposal in a slightly different context.

convention should be.⁴¹ So, while intuitively we take moral inquiry to be a process of *discovery*, it is—by the conventionalist’s account—really a process of *creation*. By contrast, my subjectivism does not have this counter-intuitive result. Moral inquiry is a process of *discovery*—we’re trying to find the judgments we’d stand by.

* * *

Taken together, the argument of the last several sections provides a powerful demonstration of the ability of my sophisticated subjectivism to account for the central features of our moral discourse and practice. Moreover, we also have a viable descriptivist account of moral error—that is, we have an account of moral facts that can meet the two tests developed in chapters 2 and 3. When combined with the critical arguments of the previous chapters, we get a compelling case for an un(der)appreciated account of moral facts.

⁴¹ See, for instance, Harman & Thomson 1996: 29-31 and Wong 2006: chap. 9.

References

- Adams, Robert. (1973) "A Modified Divine Command Theory of Ethical Wrongness." In *Religion and Morality: A Collection of Essays*. Eds. Gene Outka and John P. Reeder. New York: Doubleday.
- . (1987) "Divine Commands and the Social Nature of Obligation." *Faith and Philosophy*, IV: 262-275.
- Aronson, Elliot. (1999) *The Social Animal* 8th edition. New York: Worth Publishers/Freeman.
- Ayer, A. J. (1936/1946) *Language, Truth, and Logic*. 2nd edition. New York: Dover.
- Blackburn, Simon. (1984) *Spreading the Word*. Oxford: Oxford University Press.
- . (1987) "Morals and Modals." In Blackburn (1993).
- . (1988) "Attitudes and Contents." In Blackburn (1993).
- . (1993) *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- . (1998) *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Clarendon.
- Bloomfield, Paul. (2001) *Moral Reality*. Oxford: Oxford University Press.
- Boghossian, Paul. (1989) "The Status of Content." *Philosophical Review*. 99.2: 157-84.
- Boyd, Richard. (1988) "How to be a Moral Realist." In *Essays on Moral Realism*. Ed. G. Sayre-McCord. Ithaca: Cornell University Press.
- Brandom, Robert. (1994) *Making it Explicit*. Cambridge, MA: Harvard University Press.
- . (2001) *Articulating Reasons*. Cambridge, MA: Harvard University Press.
- Brandt, Richard. (1979) *A Theory of the Good and the Right*. Oxford: Clarendon.
- Brink, David. (1984) "Moral Realism and the Sceptical Arguments from Disagreement and Queerness." *Australasian Journal of Philosophy* 62: 111-25.
- . (1989) *Moral Realism and the Foundation of Ethics*. Cambridge: Cambridge University Press.

- . (2006) "The Significance of Desire." In *Oxford Studies in Metaethics, Vol. 3*. Ed. R. Shafer-Landau. Oxford: Oxford University Press, 5-46.
- Burge, Tyler. (1996) "Our Entitlement to Self-Knowledge." *Proceedings of the Aristotelian Society* 96: 91-116.
- Cheney, Dorthy and Robert Seyfarth. (2008) *Baboon Metaphysics: The Evolution of a Social Mind*. Chicago: University of Chicago Press.
- Churchland, Paul. (1981) "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy*, 78(2).
- Cohen, Jonathan. (2009) *The Red and the Real*. Oxford: Oxford University Press.
- Copp David. (1991) "Moral Realism: Facts and Norms." *Ethics*. 101: 610-24.
- Dancy, Jonathan. (1993) *Moral Reasons*. Oxford: Blackwell.
- D'Arms, Justin. (2006) "Two Arguments for Sentimentalism." *Philosophical Issues* 15: 1-21.
- D'Arms, Justin and Daniel Jacobson. (2000) "Sentiment and Value." *Ethics* 110: 722-48.
- . (2006) "Anthropocentric Constraints on Human Value." In *Oxford Studies in Metaethics, Vol. 1*. Ed. R. Shafer-Landau. Oxford: Oxford University Press.
- Darwall, Stephen. (2006) *The Second Person Standpoint*. Cambridge, MA: Harvard University Press.
- Darwall, S., A. Gibbard, P. Railton (1992) "Toward Fin de Siecle Ethics: Some Trends." *Philosophic Review* 101: 115-89.
- Davidson, Donald. (1984) *Inquiries into Truth and Interpretation*. Oxford: Clarendon.
- Dreier, James. (1999) "Transforming Expressivism." *Nous* 33.4: 558-72.
- Dummett, Michael. (1991) *Logical Basis of Metaphysics*. Cambridge, MA: Harvard University Press.
- Etchemendy, John. (1990) *The Concept of Logical Consequence*. Cambridge, MA: Harvard University Press.
- Feldman, Fred. (1978) *Introductory Ethics*. New York: Prentice-Hall.

- . (2002) "The Good Life: A Defense of Attitudinal Hedonism," *Philosophy and Phenomenological Research* LXV(3): 604-628.
- . (2004) *Pleasure and the Good Life*. Oxford: Clarendon.
- Firth, Roderick. (1952) "Ethical Absolutism and the Ideal Observer." *Philosophy and Phenomenological Research* 12(3): 317-45.
- Foot, Philippa. (1995) "Does Moral Subjectivism Rest on a Mistake?" *Oxford Journal of Legal Studies* 15: 1-14.
- Frank, Robert. (1988) *Passions within Reason*. New York: Norton.
- Frankena, W. K. (1939) "The Naturalistic Fallacy." *Mind* 48: 464-77.
- Frege, Gottlob. (1893/1903) *Grundgesetze der Arithmetik. Jena: Verlag Hermann Pohle, Band I, II*. Partial translation. M. Furth. Berkeley: University of California Press, 1964.
- Gauthier, David. (1985) *Morals by Agreement*. Oxford: Oxford University Press.
- Gewirth, Alan. (1980) *Reason and Morality*. Chicago: University of Chicago Press.
- Gibbard, Alan. (1990) *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
- . (2003) *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Gilbert, Margaret. (1989) *On Social Facts*. New York: Routledge.
- Gill, Michael and Shaun Nichols. (2008) "Sentimentalist Pluralism: Moral Psychology and Philosophical Ethics." *Philosophical Issues* 18, 143-163.
- Grice, Paul. (1989) *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Goodwin, G.P. and J. M. Darley. (2008) "The Psychology of Meta-Ethics: Exploring Objectivism." *Cognition* 106(3): 1339-66.
- Hacking, Ian. (1975) "What is Logic?" *Journal of Philosophy* 76(6): 285-319.
- Hale, Robert. (1993) "Can There be a Logic of Attitudes?" In *Reality, Representation, and Projection*. Eds. J. Haldane and C. Wright. Oxford: Oxford University Press.
- Haidt, Jonathan. (2006) *The Happiness Hypothesis*. New York: Basic Books.

- . (2003) "The Moral Emotions." In *Handbook of Affective Sciences*. Eds. R. J. Davidson, K. R. Scherer, & H. H. Goldsmith. Oxford: Oxford University Press, 852-870.
- . (2001) "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108(4): 814-34.
- Haidt, Jonathan et. al. (1993) "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?" *Journal of Personality and Social Psychology*, 65: 613-628.
- Haidt, Jonathan, and C. Joseph. (2004) "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues." *Daedalus*, 55-66.
- Haidt, Jonathan, and S. Kesebir. (2009) "Morality." In *The Handbook of Social Psychology*. 5th edition. Eds. S.T. Fiske and D. Gilbert. New York: Wiley.
- Hampton, Jean. (1993) "Selflessness and Loss of Self." *Social Philosophy and Policy* 10(1): 135-65.
- Harman, Gilbert. (1975) "Moral Relativism Defended." *Philosophical Review* 84: 3-22.
- . (1977) *The Nature of Morality*. Oxford: Oxford University Press.
- . (2005) "Moral Particularism and Transduction." *Philosophical Perspectives* 15: 44-55.
- Harman, Gilbert and J. J. Thompson. (1996) *Moral Relativism and Moral Objectivity*. Oxford: Blackwell.
- Hare, R. M. (1952) *The Language of Morals*. Oxford: Clarendon Press.
- . (1981) *Moral Thinking*. Oxford: Oxford University Press.
- Horgan, Terence. (1986) "Truth and Ontology." *Philosophical Papers* 15, 1-21.
- Horgan, Terence and Mark Timmons. (2006) "Cognitivist Expressivism." In *Metaethics After Moore*. Eds. T. Horgan and M. Timmons. Oxford: Oxford University Press.
- Horwich, Paul. (1994) "The Essence of Expressivism." *Analysis* 45.
- . (1998) *Truth*. 2nd edition. Oxford: Oxford University Press.

- Horyn, Cathy. (2009) "Prada: Gone Fishin'." *The New York Times*.
<http://runway.blogs.nytimes.com/2009/03/01/prada-gone-fishin/?apage=3#comments>.
- Hume, David. (1888) *Treatise on Human Nature*. Ed. L.A. Selby-Biggs. Oxford: Oxford University Press.
- Jacobson, Daniel. (2005) "Seeing by Feeling: Virtues, Skills, and Moral Perception," *Ethical Theory and Moral Practice* 8: 387-409.
- Jackson, Frank. (1998) *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- Johnston, Mark. (1989) "Dispositional Theories of Value." *Proceedings of the Aristotelian Society*. LXIII, supp.: 139-74.
- Kamm, Francis. (1993) *Morality and Mortality, Vol. 1*. Oxford: Oxford University Press.
- Kant, Immanuel. (1785) *Groundwork of the Metaphysics of Morals*. Ed. M. Gregor. Cambridge: Cambridge University Press.
- Kavka, Gregory. (1983) "The Toxin Puzzle." *Analysis* 43(1): 33-6.
- Kelly, Daniel. (ms.) "The Ethics of Disgust." Unpublished manuscript.
- Korsgaard, Christine. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.
- . (2009) *Self-Constitution*. Oxford: Oxford University Press.
- Kurth, Charlie. (2003) *Agency and Rationality: Toward a Complete Account of Self-Knowledge*. Masters Thesis. Arizona State University: Tempe, AZ.
- . (2007) "Comments on 'What's Right about the Open Question Argument.'" APA–Pacific Division, San Francisco, CA.
- . (ms. 1) "Making Sense of Moral Motivation." Unpublished manuscript.
- . (ms. 2) "Fashion Models and Moral Realists." Unpublished manuscript.
- . (ms. 3) "In Defense of Assurance-Based Accounts of Promises." Unpublished manuscript.
- Leiter, Brian. (2001) "Moral Facts and Best Explanations." *Social Philosophy & Policy* 18: 79-101.

- Lewis, David. (1969) *Convention*. Cambridge, MA: Harvard University Press.
- . (1979) "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8: 339-59.
- . (1989) "Dispositional Theories of Value." *Proceedings of the Aristotelian Society*. LXIII, supp.: 113-37.
- Lillehammer, Hallvard. (2007) *Companions in Guilt*. New York: Palgrave MacMillan.
- Linsky, Bernard and Edward Zalta. (1995) "Naturalize Platonism versus Platonism Naturalized." *Journal of Philosophy* 92(10): 525-555.
- Mackie, John. (1977) *Ethics: Inventing Right and Wrong*. New York: Penguin.
- Maddy, Penelope. (1990) *Realism in Mathematics*. New York: Oxford University Press.
- MacFarlane, John. (2005) "Logical Constants." *The Stanford Encyclopedia of Philosophy*. Spring 2009 edition. Ed. E. N. Zalta.
(<http://plato.stanford.edu/archives/spr2009/entries/logical-constants>).
- McDowell, John. (1985) "Values as Secondary Qualities." In *Morality and Objectivity: A Tribute to J.L. Mackie*. Ed. T. Honderich. London: Routledge & Kegan Paul.
- . (1997) "Brandom on Representation and Inference." *Philosophy and Phenomenological Research* 57(1): 157-62.
- McGee, Vann. (1996) "Logical Operations." *Journal of Philosophical Logic* 25: 567-580.
- Miller, Alexander. (2003) *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press.
- Millikan, Ruth. (2005) *Language: A Biological Model*. Oxford: Clarendon Press.
- Moore, G. E. (1903) *Principia Ethica*. Cambridge: Cambridge University Press.
- Moran, Richard. 2001. *Authority and Estrangement*. Princeton: Princeton University Press.
- Nagel, Thomas. (1970) *The Possibility of Altruism*. Princeton: Princeton University Press.

- . (1979) "The Fragmentation of Value." In *Mortal Questions*. Cambridge: Cambridge University Press.
- Nichols, Shaun. (2004) *Sentimental Rules*. New York: Oxford University Press.
- Parfit, Derek. (1984) *Reasons and Persons*. Cambridge, MA: Harvard University Press.
- Peacocke, Christopher. (1976) "What is a Logical Constant?" *Journal of Philosophy* 73(9): 221-40.
- Quine, W.V.O. (1960) *Word and Object*. Cambridge, MA: MIT Press.
- . (1980) "Grammar, Truth, and Logic." In *Philosophy and Grammar*. Eds. S. Kanger and S. Öman. Dordrecht: Reidel, 17-28.
- Railton, Peter. (1986a) "Moral Realism." In Railton (2003).
- . (1986b) "Facts and Values." In Railton (2003).
- . (1989) "Naturalism and Prescriptivity." *Social Philosophy & Policy* 7(1): 151-74.
- . (1997) "Aesthetic Value, Moral Value, and the Ambitions of Naturalism." In Railton (2003).
- . (1998) "Red, Bitter, Good." In Railton (2003).
- . (2003) *Facts, Values, and Norms*. Cambridge: Cambridge University Press.
- Robbins, Philip. (2004) "To Structure, or Not to Structure?" *Synthese* 139: 55-80.
- Rosati, Connie. (1995) "Persons, Perspectives, and Full Information Accounts of the Good." *Ethics* 105: 296-325.
- Rosen, Gideon. (1997) "Who Makes the Rules Around Here?" *Philosophy and Phenomenological Research* 57(1): 163-72.
- . (1998) "Blackburn's Essays in Quasi-Realism." *Nous* 32(3): 386-405.
- Ross, H. D. (1930) *The Right and the Good*. Oxford: Oxford University Press.
- Safer-Landau, Russell. (2003) *Moral Realism: A Defense*. Oxford: Clarendon Press.
- Sainsbury, Mark. (1991) *Logical Forms*. Cambridge, MA: Blackwell.

- Sayre-McCord, Geoffrey. (1988) "Moral Theory and Explanatory Impotence." In *Essays on Moral Realism*. Ed. G. Sayre-McCord. Ithaca: Cornell University Press, 256-81.
- . (ms.) "A Moral Argument Against Moral Dilemmas." Unpublished manuscript.
- Scanlon, T. M. (1998) *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Schroeder, Mark. (2008) *Being For*. Oxford: Oxford University Press.
- . (2009) "Hybrid Expressivism: Virtues and Vices." *Ethics* 119: 257-309.
- Schueler, G. (1988) "Modus Ponens and Moral Realism." *Ethics* 98(3): 492-500.
- Sher, Gila. (1991) *The Bounds of Logic: A Generalized Viewpoint*. Cambridge, MA: MIT Press.
- . (1999) "On the Possibility of a Substantive Theory of Truth." *Synthese* 117.
- Shiffrin, Seana. (1999) "Moral Overridingness and Moral Subjectivism." *Ethics* 109: 772-94.
- Shoemaker, Sydney. (1984) *Identity, Cause, and Mind*. New York: Cambridge University Press.
- Sidgwick, Henry. (1877) *The Method of Ethics*. London: Macmillan.
- Singer, Peter. (1972) "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1: 229-43.
- Smith, Michael. (1989) "Dispositional Theories of Value." *Proceedings of the Aristotelian Society*. LXIII, supp.: 89-111.
- . (1995) *The Moral Problem*. Oxford: Blackwell.
- Soames, Scott. (2005) *Philosophical Analysis in the Twentieth Century*. Princeton: Princeton University Press.
- Sobel, David. (1994) "Full Information Accounts of Well-Being." *Ethics* 104: 784-810.
- Sober, Elliot. (1982) "Realism and Independence." *Nous* XVI(3): 369-86.
- Stevenson, C. L. (1937) "The Emotive Meaning of Ethical Terms." *Mind* 46: 14-31.

- Sturgeon, Nicholas. (1985a) "Gibbard on Moral Judgment and Norms." *Ethics* 96(1): 22-33.
- . (1985b) "Moral Explanations." In *Morality, Reason, and Truth*. Eds. D. Copp and D. Zimmerman. Totowa, NJ: Rowman and Allanheld, 49-78.
- . (1986) "What Difference Does it Make Whether Moral Realism Is True." *Southern Journal of Philosophy*. 24, supp.: 115-42.
- . (1994) "Moral Disagreement and Moral Relativism." *Social Policy and Philosophy* 11(1): 80-115.
- . (1995) "Critical Study: Wise Choices, Apt Feelings." *Nous*. 29(3): 402-24.
- Sudgen, Robert. (1986/2001) *The Economics of Rights, Co-Operation, and Welfare*. Oxford: Basil Blackwell.
- Stich, Stephen. (1992) "What is a Theory of Mental Representation?" *Mind* 101: 243-61.
- Tarski, Alfred. (1935) "The Establishment of Scientific Semantics." In Tarski (1983).
- . (1983) *Logic, Semantics, Mathematics*. 2nd edition. Indianapolis, IN: Hackett.
- . (1986) "What are Logical Notions?" *History and Philosophy of Logic* 7: 143-154.
- Timmons, Mark. (1999) *Morality without Foundations*. Oxford: Oxford University Press.
- . (1990) "On the Epistemic Status of Considered Moral Judgments." *Southern Journal of Philosophy* 29, supp.: 97-129.
- Thomson, J.J. (1990) *The Realm of Rights*. Cambridge, MA: Harvard University Press.
- Unwin, Nicholas. (2001) "Norms and Negation: A Problem for Gibbard's Logic." *Philosophical Quarterly* 51: 60-75.
- van Roojen, Mark. (1996) "Expressivism and Irrationality." *Philosophical Review* 105: 311-35.
- Velleman, J. David. (1988) "Brandt's Definition of the Good." *Philosophical Review* 97: 353-71.
- Wedgwood, Ralph. (2007) *The Nature of Normativity*. Oxford: Oxford University Press.

- Wiggins, David. (1987) "A Sensible Subjectivism." In *Needs, Value, and Truth*. Oxford: Basil Blackwell.
- Wittgenstein, Ludwig. (1921) *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul.
- Wong, David. (2006) *Natural Moralities*. Oxford: Oxford University Press.
- . (1984) *Moral Relativity*. Berkeley, CA: University of California Press.
- Wright, Crispin. (1988) "Realism, Antirealism, Irrealism, Quasi-Realism." In *Midwest Studies in Philosophy, XII*. Eds. P. French, T. Uehling, and H. Wettstein. Minneapolis: University of Minnesota Press.
- . (1992) *Truth and Objectivity*. Cambridge, MA: Harvard University Press.
- Yablo, Stephen. (1995) "Singling Out the Properties." *Philosophical Perspectives* 9: 477-502.