

NON-COGNITIVISM, INTERNALISM, AND THE FREGE-GEACH PROBLEM

A Dissertation
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree

Doctor of Philosophy

by
Jason Berntsen

Dr. Robert Johnson, Dissertation Supervisor

August 2007

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

NON-COGNITIVISM, INTERNALISM, AND THE FREGE-GEACH PROBLEM

presented by Jason Berntsen,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Robert Johnson

Professor Peter Vallentyne

Professor Andrew Melnyk

Professor Brian Kierland

Professor Matthew McGrath

Professor Ines Segert

To Steph, Mom, Dad, Jason, and Darin

ACKNOWLEDGMENTS

I received tremendous support while writing this dissertation. My advisor, Robert Johnson, and the other members of my dissertation committee, Peter Vallentyne, Andrew Melnyk, Brian Kierland, and Matthew McGrath are exceptional philosophers and academic role models. They all provided generous feedback and guidance throughout each stage of the writing process. Furthermore, their encouragement and faith in my philosophical abilities helped me to overcome many moments of frustration and despair. I also benefited from MU's Kline Chair Workshops which gave me the opportunity to discuss metaethics with such excellent philosophers as David Brink, John Broome, James Dreier, Geoffrey Sayre-McCord, Mark Schroeder, and Michael Smith. Informal conversations with other MU graduate students were also extremely helpful. I especially benefited from my many conversations with Jared Bates, Darin Finke, Jason Hedderman, Eric Heidenreich, Jim McBain, Justin McBrayer, Eric Roark, Alan Tomhave, and Yeongseo Yeo, conversations which I count as some of the most rewarding and pleasant experiences of my life. Mary Jane Berntsen and Stephanie Craft both graciously proofread the penultimate draft. Stephanie also provided valuable stylistic suggestions on earlier drafts. Finally, it is impossible to convey the importance of all of the love and support that I received from Stephanie Craft, Mary Jane Berntsen, Kenny Berntsen, Jason Hedderman, and Darin Finke. As my mother would say, I am lucky-blessed.

TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

Chapter

1. INTRODUCTION 1

2. TWO KINDS OF NON-COGNITIVISM 5

Introduction 5

Two Philosophical Implications of the Distinction 6

The Objection from Amoralism 6

The “Express” Relation 10

Joyce and Mental State Non-cognitivism 12

Arguing from Motivation Internalism to Mental State
Non-cognitivism 12

Mental State Non-cognitivism and Hurried Moral Judgments 16

An Empirical Debate? 19

Conclusion 23

3. CAN INTERNALISM EXPLAIN THE PRACTICAL CHARACTER OF
MORALITY? 25

Introduction 25

The Practical Character of Morality 26

Two Weaker Versions of Internalism 29

Very Strong Internalism and Overriding Defeasible Internalism 38

Conclusion 49

4. THE FREGE-GEACH PROBLEM 50

Introduction	50
Non-cognitivism	51
The Simple Argument	53
The Equivocation Argument	55
The Standard Validity Argument	61
The Shared Principle Argument	66
Conclusion	67
5. WHY EMOTIVISTS SHOULDN'T LOVE INCONSISTENCY	69
Introduction	69
The Problem of Reasoning	69
Björnsson's Solution	75
Two Problems with Björnsson's Solution	82
Conclusion	86
WORKS CITED	88
VITA	91

Chapter One

Introduction

Some people are morally opposed to capital punishment. But what does such opposition consist in? Is it a matter of having a certain *belief*, of thinking that things are a certain way? Consider a paradigmatic belief: I believe that Austin is the capital of Texas. Is the state of mind of being morally opposed to capital punishment the same *kind* of state of mind as my belief about Austin? Suppose it is not. In that case, what kind of state of mind *is* the view that capital punishment is morally wrong? These are, of course, instances of two general questions: Are moral views beliefs? If they are not beliefs, what *are* they?

Non-cognitivists about moral thought give a negative answer to the first question. According to non-cognitivists, although we speak colloquially about “moral beliefs,” strictly speaking there are no such things. To have a moral view is not, on this view, a matter of thinking that things are a certain way. What, then, are moral views according to non-cognitivists? Non-cognitivists claim that to have a moral view about something is just to have an attitude of approval or disapproval toward that thing. Suppose that Jones is one of those who is morally opposed to capital punishment: He thinks that capital punishment is always morally wrong. Smith has a different view: He thinks that capital punishment is often morally required. Non-cognitivism implies that the difference between Jones and Smith is *not* a difference in belief, not, that is, the kind of difference that exists between me and someone who believes that, say, San Antonio is the capital of

Texas. Rather, the difference is merely one of attitude: Jones disapproves of all instances of capital punishment; Smith approves of some instances of it.¹

I have thus far presented non-cognitivism as a view about moral *thought*. It is also often cast as a view about moral *discourse*. We use language to describe the world, to state (what we take to be) facts. That, for example, is the typical function of sentences like “the door is closed.” But stating facts is, of course, not the only thing we use language for.² Consider the expression “Ouch!” We ordinarily use that expression not to describe how things are but to give voice to pain, to *express* pain, where *expressing pain* is distinct from *saying that one is in pain*. (Compare uttering “Ouch!” with uttering “I am in pain.”) Corresponding to the above two questions about moral thought are the following two questions about moral discourse: Is the function of moral language to state facts? If it is not, what *is* the function of moral language. As a view about moral discourse, non-cognitivism holds that, despite appearances, moral language is like “Ouch!” and unlike “the door is closed” in that it does not play a fact-stating role. According to non-cognitivists, just as the function of “Ouch!” is to express pain, the function of moral language is to express attitudes of approval and disapproval.³

¹ I am here ignoring the distinction between thinking that an action is morally permissible and thinking that an action is morally required.

² See Charles Stevenson, “The Emotive Meaning of Ethical Terms,” in *Moral Discourse and Practice: Some Philosophical Approaches*, ed. Stephen Darwall, Allan Gibbard, and Peter Railton (Oxford: Oxford University Press, 1997), 75-76.

³ Alexander Miller, *An Introduction to Contemporary Metaethics* (Cambridge, England: Polity Press, 2003), 37. Prominent non-cognitivists include A.J. Ayer, Charles Stevenson, Simon Blackburn, and Allan Gibbard. See A.J. Ayer, *Language, Truth, and Logic*, 2d ed. (n.p.: Gollantz, 1946; reprint, New York: Dover Publications, 1952); Stevenson, “Emotive Meaning”; Simon Blackburn, *Spreading the Word: Groundings in the Philosophy of Language* (Oxford: Oxford University Press, 1984), 145-223; Simon Blackburn, *Ruling Passions: A Theory of Practical Reasoning* (Oxford: Oxford University Press, 1998); Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge, Massachusetts: Harvard University Press, 1990); Allan Gibbard, *Thinking How to Live* (Cambridge, Massachusetts: Harvard University Press, 2003).

This dissertation is about the prospects for non-cognitivism. Chapter Two focuses on the distinction between non-cognitivism about moral thought and non-cognitivism about moral discourse. This distinction is often overlooked. I show, however, that it has at least two important philosophical implications. First, non-cognitivists about moral discourse have a response to a standard objection to non-cognitivism that is not available to non-cognitivists about moral thought. Second, non-cognitivists about moral thought sidestep a thorny issue facing non-cognitivists about moral discourse: the issue of what sense of “expression” is at work in the claim that moral sentences express attitudes of approval and disapproval. I also show that Richard Joyce’s recent attempt to show that the “real” philosophical debate is over the truth of non-cognitivism about moral discourse fails.⁴

The standard objection to non-cognitivism that I discuss in Chapter Two appeals to motivation internalism. According to this view, there is a necessary connection between having a moral view and having a corresponding motivation.⁵ Motivation internalism also figures in a standard argument *for* non-cognitivism. Indeed, one of Joyce’s reasons for thinking that the “real” philosophical debate concerns non-cognitivism about moral discourse is that, as Joyce sees things, the argument for non-cognitivism that is based on motivation internalism does not apply to non-cognitivism about moral thought.⁶ Motivation internalism thus features prominently in Chapter Two. I continue to look at motivation internalism in Chapter 3. Proponents of motivation internalism sometimes argue that motivation internalism is the best explanation of the

⁴ Richard Joyce, “Expressivism and Motivation Internalism,” *Analysis* 62, (October 2002): 341-343.

⁵ Miller, *Introduction to Metaethics*. 7.

⁶ *Ibid.*, 342-343

practical character of morality.⁷ There are, however, different versions of motivation internalism, and I show that two popular versions cannot in fact explain why moral matters are practical matters.

Chapters Four and Five both deal with what is widely considered to be one of the most pressing objections to non-cognitivism, the so-called Frege-Geach problem. In Chapter Four I argue that there is, in fact, no such thing as *the* Frege-Geach problem and distinguish four distinct arguments that have been discussed under that heading. I show that while none of these arguments are conclusive as they stand, they each provide the non-cognitivist with a distinct challenge. I briefly sketch ways in which non-cognitivists can attempt to answer each challenge. In Chapter Five I present a recently proposed solution to “the problem of reasoning,” a Frege-Geach style argument against non-cognitivism, and argue that that solution does not work.

⁷ See, for example, Michael Smith, *The Moral Problem* (Oxford: Blackwell Publishers, 1994), 71-84.

Chapter Two

Two Kinds of Non-cognitivism

Introduction

We can distinguish two related metaethical views that have both been discussed under the labels “non-cognitivism” and “expressivism.” The first is a view about what certain utterances or speech acts express. According to this view moral utterances, utterances like “abortion is wrong,” express not beliefs but rather “desires, emotions, or pro/con attitudes.”⁸ The second view is a view about what certain mental states are. The view holds that the mental states that we refer to when we say things like “Geri believes that abortion is wrong” are not *really* beliefs at all but are actually desires, emotions, or pro/con attitudes.” These views are obviously related. Indeed, they may even appear at first glance to be mere notational variants on one another. They are, in fact, often conflated in the literature. This conflation is, I think, encouraged by the fact that both views are usually formulated as views about “moral judgments,” a term that is sometimes used to refer to a kind of utterance and sometimes used to refer to a kind of a mental state. Some philosophers do clearly mark this ambiguity of “moral judgment” and the distinction between non-cognitivism about moral utterances and non-cognitivism about mental states.⁹ However, one of these philosophers, Richard Joyce, introduces the distinction only to warn against construing “the” cognitivist/non-cognitivist debate in

⁸ Joyce, “Expressivism and Internalism,” 337.

⁹ Jonathan Dancy and Russ Shafer-Landau, for example, conflate the two senses of “moral judgment” and the two corresponding types of non-cognitivism. See Jonathan Dancy, *Moral Reasons* (Oxford: Blackwell Publishers, 1993), 3; Russ Shafer-Landau, *Moral Realism: A Defence* (Oxford: Oxford University Press, 2003), 18. Richard Joyce and James Dreier are examples of philosophers who mark the distinction between the two senses of “moral judgment.” See Joyce, “Expressivism and Internalism,” 342; James Dreier, “Troubling Developments in Metaethics,” *Noûs* 36 (2002): 168.

mentalistic terms. According to Joyce, the “real” philosophical debate is over the first of the two views.¹⁰

I do two things in this chapter. First, I argue that although the distinction between these two metaethical views is frequently overlooked, it has at least two important philosophical implications. Second, I argue that Joyce’s attempt to show that the “real” debate is over the truth of the first view fails. The upshot is that (a) every non-cognitivist has to decide whether she accepts the first of the above two views, the second, or both, and (b) Joyce has not given us any reason to focus all of our energies on the first view.

Two Philosophical Implications of the Distinction

It will be helpful to stipulate clear labels for the two views I have just distinguished. Since the first view concerns moral *utterances* or *speech acts*, let’s call it “linguistic non-cognitivism.” And since the second view is about the kind of mental state that we refer to when we say things like “Geri believes that abortion is wrong,” we can call it “mental state non-cognitivism.” It is unfortunate that both of these views are often conflated under the label “non-cognitivism” or “expressivism,” for as I will now show, there are at least two important philosophical differences between linguistic non-cognitivism and mental state non-cognitivism.

The Objection from Amoralism

The first difference is that linguistic non-cognitivists have a reply to a standard objection to non-cognitivism that is not available to mental state non-cognitivists. The objection relies on the premise that non-cognitivism implies motivation internalism. Like

¹⁰ This at least is how I interpret Joyce. Joyce argues that a standard argument for non-cognitivism does not apply to mental state non-cognitivism. He also identifies considerations that he thinks “cast the usefulness of the mentalistic version of ... [non-cognitivism] into doubt.” Finally, he argues that construing the debate between cognitivists and non-cognitivists in mentalistic terms has “a very unwelcome result.” Joyce, “Expressivism and Internalism,” 342-343.

non-cognitivism, motivation internalism is usually stated as a thesis about “moral judgments.” Joyce, who presents the reply I have in mind, gives a standard formulation: motivation internalism is the view that “[i]t is necessary and a priori that for any x and any y : if x judges that some available action y is morally right (good, obligatory, etc.), then x will have some (defeasible) motivation in favour of performing y .”¹¹ The objection runs as follows. Motivation internalism implies that it is impossible for someone to judge that it is right to act in a certain way and not have any motivation to act in that way. But that *is* possible. Thus, motivation internalism is false. But non-cognitivism implies motivation internalism. Thus, non-cognitivism is false. Call this “the objection from amorality.”¹²

Joyce responds to this objection by attacking the premise that non-cognitivism implies motivation internalism. His strategy is to undermine what he takes to be the principal reason for thinking that non-cognitivism *does* imply motivation internalism. Why might someone think that non-cognitivism implies motivation internalism? Joyce’s target is the following line of reasoning. If non-cognitivism is true, then the judgment that an action is right expresses a pro-attitude toward that action. And if the judgment that an action is right expresses a pro-attitude toward that action, then an agent x judges that some action y is right only if x has a pro-attitude toward y . But if x has a pro-attitude toward y , then x has some motivation to perform y . Therefore, if non-cognitivism is true, then x judges that y is morally right only if x has some motivation to perform y .

The problem with this reasoning, according to Joyce, is that it does not follow from the claim that the judgment that an action is right *expresses* a pro-attitude toward

¹¹.Joyce, “Expressivism and Internalism,” 337.

¹² Ibid., 339. See also David Brink, *Moral Realism and the Foundations of Ethics*, (Cambridge, England: Cambridge University Press, 1989), 46.

that action that anyone who makes the judgment *has* the relevant pro-attitude. That does not follow, according to Joyce, because moral judgments are utterances, and, quite generally, that a kind of utterance expresses a certain mental state does not imply that anyone who makes the utterance is in the relevant mental state. This last is true, Joyce argues, because whether an utterance expresses a mental state (in the relevant sense of “express”) is determined by linguistic convention.¹³ Joyce cites apologies, promises, and assertions. An utterance of “I am sorry” made under certain conditions determined by linguistic convention expresses regret, and this is so even if the utterer in fact feels no regret. If someone utters “I am sorry” (under the right conditions) and feels no regret, then that person may be offering an insincere apology, but insincere apologies are still apologies; and part of what one does when one apologizes is express regret.¹⁴ (Someone who fails to express regret fails to apologize, and thus fails to apologize insincerely.) An utterance of “I promise to feed the dog while you are away” made under certain conditions determined by linguistic convention expresses an intention or commitment to feed the dog, and this is so even if the utterer in fact has no such intention. Someone who says this (under the right conditions) may be making a false promise, but false promises are still promises; and part of what one does when one makes a promise is express a certain intention or commitment. (Someone who fails to express an intention or commitment fails to make a promise, and thus fails to make a false promise.) An utterance of “grass is green” made under certain conditions determined by linguistic

¹³ Joyce, “Expressivism and Internalism,” 338. Note that Joyce thus denies that the relevant sense of “express” is causal. See *ibid.*, 343.

¹⁴ I say that such a person *may* be making a false promise because, as Joyce notes, it is not obvious that sincerely performing a speech act requires being in the mental state that that speech act expresses. *Ibid.*, 341-342. See also Michael Ridge, “Sincerity and Expressivism,” *Philosophical Studies* 131 (November 2006).

convention expresses the belief that grass is green even if the utterer does not in fact have that belief. The utterer who says this (under the right conditions) without having the belief is asserting something that she does not believe, but she is still making an assertion; and part of what one does when one makes an assertion is express a belief. (Someone who fails to express a belief fails to assert something, and thus fails to assert something that she does not believe.)¹⁵ Likewise, Joyce concludes, it may be that moral judgments express attitudes and yet be possible for someone to make a moral judgment without having the relevant attitude.¹⁶

It should be clear that this response to the objection from amoralism works only if moral judgments are treated as utterances or speech acts and non-cognitivism is treated as linguistic non-cognitivism.¹⁷ If we treat moral judgments as mental states and non-cognitivism as mental state non-cognitivism, then the objection's premise that non-cognitivism implies motivation internalism does not rely on moving from the idea that an utterance expresses a certain mental state to the idea that anyone who performs that utterance is in the relevant mental state.¹⁸ So Joyce's charge that that move is illegitimate

¹⁵ Mark Schroeder also defends this point. Mark Schroeder, "Expression for Expressivists," forthcoming in *Philosophy and Phenomenological Research*.

¹⁶ Joyce, "Expressivism and Internalism," 337-339.

¹⁷ One might think that the reply does not work even with respect to linguistic non-cognitivism on the grounds that linguistic non-cognitivism, properly understood, is not a thesis about *all* moral speech acts but only about *sincere* moral speech acts. The idea would be that even if Joyce is right that it does not follow from the claim that a moral speech act expresses an attitude that the speaker has that attitude, that the speaker has the attitude *does* follow from the claim that the speaker *sincerely* performs a speech act that expresses that attitude. Joyce anticipates this objection. He replies by denying the initially plausible view—defended by John Searle—that a speech act is sincere only when the speaker is in the mental state that the speech act expresses. *Ibid.*, 341-342; John Searle, *Speech Acts: An Essay in the Philosophy of Language* (New York: Cambridge University Press, 1969). Michael Ridge also denies this view, but for different reasons. Ridge, "Sincerity and Expressivism."

¹⁸ The argument for the claim that non-cognitivism implies motivation internalism would then simply be the following. If non-cognitivism is true, then *x* judges that some action *y* is right only if *x* has a pro-attitude toward *y*—since according to linguistic non-cognitivism, the judgment that some action *y* is right *just is* a pro-attitude toward *y*. But if *x* has a pro-attitude toward *y*, then *x* has some motivation to perform *y*. Therefore, if non-cognitivism is true, then *x* judges that *y* is morally right only if *x* has some motivation to perform *y*.

does not threaten the premise that *mental state non-cognitivism* implies motivation internalism. As a result, while Joyce's response may rescue *linguistic non-cognitivism* from the objection from amoralism, it does nothing to help the *mental state non-cognitivist* answer that objection.¹⁹

The "Express" Relation

Another issue for which the distinction between linguistic and mental state non-cognitivism matters concerns the "express" relation. We just saw that Joyce's response to the objection from amoralism relies on a general fact about expression, namely that an utterance can express a mental state that the speaker is not in. The idea that utterances express mental states is, of course, at the heart of linguistic non-cognitivism. According to that view, moral utterances express desires, emotions, or pro/con attitudes (attitudes, for short). But what exactly is this "express" relation?

As Joyce and Mark Schroeder have pointed out, very little has been said on this question.²⁰ Schroeder suggests a reason for this. The key idea behind linguistic non-cognitivism, according to Schroeder, is that moral utterances stand in *the same* relation to attitudes that ordinary, descriptive utterances stand in to beliefs, the same relation that, for example, utterances of "grass is green" stand in to the belief that grass is green.²¹ Now assuming that there is some intuitive sense in which ordinary, descriptive utterances express beliefs, it follows from the linguistic non-cognitivist's key idea that we *already*

¹⁹ This point is, I think, implicit in Mark van Roojen's response to Joyce's reply to the objection from amoralism. However, van Roojen does not disambiguate 'moral judgment' or distinguish between linguistic non-cognitivism and mental state non-cognitivism. Mark van Roojen, "Moral Cognitivism vs. Non-Cognitivism," in *The Stanford Encyclopedia of Philosophy (Winter 2005 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2005/entries/moral-cognitivism/>>.

²⁰ Joyce, "Expressivism and Internalism," 337; Schroeder, "Expression for Expressivists."

²¹ More precisely, the same relation that assertions of "grass is green" stand in to the belief that grass is green.

have an intuitive grasp on what the linguistic non-cognitivist means by “express.” So, Schroeder concludes, it is not that surprising that linguistic non-cognitivists haven’t felt the need to say much about the “express” relation.²²

Schroeder goes on to vigorously argue, however, that when we do turn to the question of what linguistic non-cognitivists can and should mean by “express” we see that they “need an account of *expression* that commits them to a quite radical view about the foundations of the semantic content of ordinary, descriptive sentences.”²³ The “radical view” that Schroeder has in mind is not important for our purposes, since the point that I want to make is just this: *mental state* non-cognitivism sidesteps any worries that might derive from questions about the “express” relation. The reason is obvious: mental state non-cognitivism does not make any claims about what moral utterances “express.” For that matter, mental non-cognitivists do not make *any claim whatever* about the nature of moral utterances.²⁴ So if, for example, one is convinced by Schroeder’s argument that the linguistic non-cognitivist’s notion of “express” commits him to a “radical view” about how ordinary, descriptive sentences like “grass is green” get their semantic content, *and* one finds this “radical view” unpalatable, then while one is thereby given a reason to reject *linguistic non-cognitivism*, one has yet to been given any reason to reject *mental state non-cognitivism*. Of course, the point is not just that mental state non-cognitivists are immune to an objection based on Schroeder’s argument. Whatever one thinks of Schroeder’s argument, one might wonder whether there is a sense of “express” that is suitable for linguistic non-cognitivism. And the point is that this is an

²² Ibid.

²³ Ibid.

²⁴ As Joyce notes, when non-cognitivism is construed mentalistically, “the whole problematic expression relation disappears.” Joyce, “Expressivism and Internalism,” 342.

issue that does not even arise for mental state expressivism. So to the extent that it is a *difficult* issue—that is, to the extent that it is difficult to determine whether there is a sense of “express” suitable for linguistic non-cognitivism—mental state non-cognitivists avoid a difficulty facing linguistic non-cognitivists.

Joyce and Mental State Non-cognitivism

So far we have seen that mental state non-cognitivism cannot avail itself of a certain response to the objection from amoralism but sidesteps another objection that can be levied against linguistic non-cognitivism. The distinction between linguistic non-cognitivism and mental state non-cognitivism thus makes a philosophical difference. As I stated earlier, Joyce is one of the few philosophers who clearly marks the distinction. However, Joyce draws the distinction only to argue that the “real” philosophical debate concerns linguistic non-cognitivism.²⁵ He presents three arguments. The first aims to show that a standard argument for non-cognitivism does not apply to mental state non-cognitivism. The second and third arguments each purport to identify a problematic implication of mental state non-cognitivism. I will now argue that all three arguments fail.

Arguing from Motivation Internalism to Mental State Non-Cognitivism

We saw that motivation internalism features in a standard objection against non-cognitivism, the objection from amoralism. But motivation internalism also appears in a standard argument *for* non-cognitivism. According to this argument, motivation internalism captures an important truth about moral judgment and non-cognitivism

²⁵ Again, Joyce does not say make this claim explicitly, but this is what I take to be the combined upshot of the three arguments described above in the main text.

provides the best explanation of this truth.²⁶ The conclusion of Joyce's first argument is that this strategy of arguing from motivation internalism to non-cognitivism is not feasible when moral judgments are treated as mental states and non-cognitivism is treated as mental state non-cognitivism. Here is Joyce's argument:

Assuming that the explication of the ... [non-cognitivist's] 'desires, emotions, or pro/con attitudes' turns out basically to mean 'motivation implicating states', then on this mentalistic construal of ... [non-cognitivism], the connections between the theory and ... [motivation internalism] will be trivial. A moral judgment (where this is some kind of mental act) will necessarily be motivating; and from the fact that a kind of judgment necessarily implies motivation it will follow that such judgments can be considered motivation-implicating. Indeed, the connections will be so trivial that arguing for either thesis by means of first establishing the other ceases to be a feasible dialectical strategy.²⁷

Joyce's argument concerns both the strategy of arguing from non-cognitivism to motivation internalism and the strategy of arguing from motivation internalism to non-cognitivism. Let's continue to focus on the latter strategy. Has Joyce shown that that strategy is not feasible when non-cognitivism is construed mentalistically? I will now argue that he has not.

Joyce's argument rests on two assumptions. The first, which Joyce makes explicit, is that when the mental state non-cognitivist says that moral judgments (which mental state non-cognitivists treat as mental acts or states) are desires, emotions, or pro/con attitudes, all he means is that moral judgments are motivation implicating states. The second, which is left implicit, is that the claim that moral judgments are desires, emotions, or pro/con attitudes is the *only* claim that is essential to mental state non-cognitivism. Let us grant, at least for the sake of argument, that if both of these

²⁶ van Roojen, "Moral Cognitivism."

²⁷ Joyce, "Expressivism and Internalism," 342-343.

assumptions are correct, Joyce's argument succeeds: Mental state non-cognitivists will have to despair of using motivation internalism to help establish their view.²⁸

On the other hand, if either of the argument's starting assumptions is false—if the mental state non-cognitivist's claim that moral judgments are desires, emotions, or pro/con attitudes is *not* equivalent to the claim that moral judgments are motivation implicating states, or if mental state non-cognitivism includes some *additional* thesis—then Joyce has not given us any reason to think that the connections between motivation internalism and mental state non-cognitivism are trivial. And if Joyce has not given us any reason to think that the connections between motivation internalism and mental state non-cognitivism are trivial, then he has not given us any reason to think that motivation internalism does not support mental state non-cognitivism. So Joyce's argument fails if either of its starting assumption is false.

The question we have to now ask, then, is whether Joyce's assumptions are both true. And the answer is no. Indeed, as I will now show, they *cannot* both be true.

Consider first linguistic non-cognitivism. It is generally agreed that one of the central disagreements between linguistic non-cognitivists and linguistic cognitivists concerns whether moral judgments (construed as utterances) express beliefs. Linguistic cognitivists think that moral judgments do express beliefs, while linguistic non-cognitivists deny that they do.²⁹ It follows that the claim that moral judgments express desires, emotions, or pro/con attitudes captures everything that is essential to linguistic

²⁸ It is actually not clear that Joyce's argument succeeds even if both of its assumptions are true. The assumptions do not render the relation between motivation internalism and mental state non-cognitivism trivial in any obvious sense, since motivational internalism is still consistent with the falsity of mental state non-cognitivism. Furthermore, someone who thinks the assumptions are both true might still think that mental state non-cognitivism is the best explanation for motivation internalism. Andrew Melnyk helped me to see this.

²⁹ Miller, "Introduction to Metaethics," 3.

non-cognitivism only if it is part of the meaning of that claim that moral judgments do *not* express beliefs. But that cannot be so if the claim that moral judgments express desires, emotions, or pro/con attitudes is equivalent to the claim that moral judgments express motivation implicating states. For someone can think that moral judgments express motivation implicating states *and* think that moral judgments express beliefs. In fact, a number of leading theorists hold just this combination of views.³⁰ It follows that the claim that moral judgments express desires, emotions, or pro/con attitudes does not capture everything that is essential to linguistic non-cognitivism if that claim is equivalent to the claim that moral judgments express motivation implicating states.

Now consider mental state non-cognitivism. Presumably, if it is an essential component of linguistic non-cognitivism that moral judgments (construed as utterances) *do not express* beliefs, then it is an essential component of mental state non-cognitivism that moral judgments (construed now as mental states) *are not* beliefs. But then reasoning that parallels that of the last paragraph shows that Joyce's two assumptions cannot both be true. Since it is an essential thesis of mental state non-cognitivism that moral judgments are not beliefs, the claim that moral judgments are desires, emotions, or pro/con attitudes captures everything essential to mental state non-cognitivism only if it is part of the meaning of that claim that moral judgments are not beliefs. But that cannot be so if the claim that moral judgments are desires, emotions, or pro/con attitudes is equivalent to the claim that moral judgments are motivation implicating states. For someone can think that moral judgments are motivation implicating states *and* think that

³⁰ Ibid., 7. Joyce himself acknowledges the possibility of holding that moral views are beliefs *and* motivation implicating. Joyce, "Expressivism and Internalism," 340.

moral judgments are beliefs.³¹ It follows that the claim that moral judgments are desires, emotions, or pro/con attitudes does not capture everything that is essential to mental state non-cognitivism if that claim is equivalent to the claim that moral judgments express beliefs. And this means that Joyce's two assumptions cannot both be true. Of course, if the two assumptions *cannot* both be true, then it follows that they *are not* both true. Yet we saw that Joyce's argument fails if the assumptions are not both true. So Joyce's argument fails. Joyce has not shown that the standard strategy of arguing from motivation internalism to non-cognitivism is not feasible when moral judgments are construed as mental states and non-cognitivism is construed as mental state non-cognitivism.

Mental State Non-cognitivism and Hurried Moral Judgments

I have done nothing to show that arguing from motivation internalism to mental state non-cognitivism *is* a feasible strategy. Rather, I have only argued that Joyce's argument for the view that it is *not* a feasible strategy fails. But even the stronger conclusion is ultimately uninteresting if we have good grounds to reject mental state non-cognitivism. And Joyce's second two arguments both purport to find problematic implications for mental state non-cognitivism.

Joyce's second argument alleges that mental state non-cognitivism implies that a moral judgment has not been made in certain cases where intuitively one has been made.

Joyce presents the following case:

Suppose that Fred and Carol have to leave the dinner party unexpectedly [J]ust prior to their hurried departure, the dinner table discussion was on the morality of Britain's keeping the Elgin marbles. As they rush to the door, Fred ... gets in his two cents worth: 'The marbles belong to the Greeks, and keeping them is wrong! Thanks and goodbye!'.³²

³¹ Michael Smith, for example, holds this view. Smith, *The Moral Problem*, 130-181.

³² Joyce, "Expressivism and Internalism," 342.

According to Joyce, “[b]y ordinary lights Fred has certainly made a public moral judgment (albeit a hurried one)”³³ But Joyce notes that things are not so simple according to mental state non-cognitivism:

According to this [mentalistic] way of thinking, in the above example of Fred’s hurried proclamation about the Elgin marbles, the . . . [non-cognitivist] would have to say that if at the moment of his departure Fred was not in some particular kind of emotive/conative state, then he did not *really* make a moral judgment at that moment. This in itself is, it seems to me, sufficiently counter-intuitive to cast the usefulness of the mentalistic version of . . . [non-cognitivism] into doubt.³⁴

I take it that the problem is supposed to be the following. Intuitively or by ordinary lights Fred made a moral judgment *even if he was not* in some particular emotive/conative state when he said, “The marbles belong to the Greeks, and keeping them is wrong,” while mental state non-cognitivism implies that Fred made a moral judgment *only if he was* in some particular emotive/conative state at that time. So mental state non-cognitivism seems to be at odds with intuition or ordinary lights in this case.

Now it is not entirely clear what conclusion Joyce draws from this apparent conflict between intuition or ordinary lights and mentalistic non-cognitivism. His explicit conclusion is that it is *dubious that it is useful* to cast non-cognitivism in mentalistic terms. He may, however, want to infer the stronger claim that mental state non-cognitivism is *just false*. Yet, as I will now show, Joyce is mistaken if he thinks that the case of Fred’s hurried proclamation about the Elgin marbles shows *either* of these things.

Note that since the term “moral judgment” is ambiguous, so are the premises of Joyce’s argument. Consider the first premise: intuitively or by ordinary lights Fred made a moral judgment even if he was not in some particular emotive/conative state when he

³³ Joyce, “Expressivism and Internalism,” 342.

³⁴ *Ibid.*

said, “The marbles belong to the Greeks, and keeping them is wrong.” This could mean *either* that (a) intuitively or by ordinary lights Fred managed to successfully perform a certain speech act even if he was not in a particular emotive/conative state at the time of his utterance *or* that (b) intuitively or by ordinary lights Fred was in a certain mental state at the time of his utterance even if he was not in a particular emotive/conative state at that time. Similarly, the second premise could mean *either* that (c) mental state cognitivism implies that Fred managed to successfully perform a certain speech act only if he was in a particular emotive/conative state at the time of utterance *or* that (d) mental state non-cognitivism implies that Fred was in a certain mental state at the time of his utterance only if he was in a particular emotive/conative state at that time. The first thing, we need to ask, then, is what “moral judgment” means in the context of this argument. Specifically, we need to ask whether it picks out a class of speech acts or a class of mental states. Let’s consider these possibilities in turn.

Suppose that in the context of this argument “moral judgment” picks out a class of speech acts. In that case, the first premise is (a). (a) is, I think, highly plausible. For it seems plausible that alongside apologizing and promising there is something like “declaring a moral view,” and that just as one can successfully apologize without feeling regret and successfully promise to do something without intending to do it, one can successfully declare a moral view without even having that view. If one can successfully declare a moral view that one does not have, then all sides can agree that one can successfully declare a moral view without being in any particular emotive/conative state. Furthermore, I see no reason to deny that Fred met the conditions for declaring a moral view. But now notice that if in the context of this argument “moral judgment” picks out a

class of speech acts, the second premise is (c), and (c) is false: Mental state non-cognitivism is silent on the conditions for successfully declaring a moral view. So if “moral judgment” is read linguistically, then at least one of the premises of Joyce’s argument is false.

Suppose, then, that in the context of this argument “moral judgment” is being used to pick out a class of mental states. In that case, the second premise is (d): mental state non-cognitivism implies that Fred was in a certain mental state (that is, the state of having the view that it is wrong to keep the Elgin marbles) at the time of his utterance only if he was in a particular emotive/conative state (that is, the state of having a con-attitude toward keeping the Elgin marbles) at that time. Since mental state non-cognitivism *identifies* the view that it is wrong to keep the Elgin marbles with a con-attitude toward keeping them, (d) is true. However, the first premise is now (b): intuitively or by ordinary lights Fred had the view that keeping the Elgin marbles is wrong at the time of his utterance even if he did not have a con-attitude toward keeping them at that time. This premise is far from obvious. More importantly, it is no less contentious than the idea that intuition or ordinary lights is at odds with mental state non-cognitivism. But that is the very thing Joyce’s two premises are supposed to establish! So, the first premise begs the question if “moral judgment” is being used mentalistically. Since “moral judgment” is being used either linguistically or mentalistically, we can conclude that Joyce’s second argument fails.

An Empirical Debate?

I turn now to Joyce’s third argument against construing non-cognitivism as mental state non-cognitivism. Joyce writes:

In any case, [construing non-cognitivism mentalistically] ... makes the whole issue an empirical one—a very unwelcome result. According to such a view, we should, at least in principle, be able to take some persons who are paradigmatic instances of ‘moral judges’, and with a PET scan watch for evidence of a ‘conative state’—say, activity in the amygdala (or whatever)—when we prompt them to think about euthanasia, or Adolph Hitler, or returning the Elgin Marbles. But this, I think most will agree, is silly; nobody imagines that the cognitivist/... [non-cognitivist] debate can be settled in such a way.³⁵

Note that this is not actually an objection to the *truth* of mental state non-cognitivism. It does not purport to show that mental state non-cognitivism is *false*.³⁶ Rather, the objection challenges a certain interpretation of the cognitivism/non-cognitivism debate. As I understand the objection, its point is that there is an important debate between cognitivists and non-cognitivists, but that debate is not over the nature of certain mental states. Joyce’s reasoning can be stated as follows. If the debate between cognitivists and non-cognitivists is a debate about the nature of certain mental states, then it can, at least in principle, be settled empirically in the way that Joyce describes. But the cognitivism/non-cognitivism debate cannot, even in principle, be settled empirically in the way that Joyce describes. Thus, the cognitivism/non-cognitivism debate is not a debate about the nature of certain mental states.

This argument is unsound. The problem is that the first premise is false. Suppose that the cognitivism/non-cognitivism debate is a debate about the nature of certain mental states. In that case, the cognitivism/non-cognitivism debate is just the debate between mental state cognitivists and mental state non-cognitivists, that is, between those who claim that the mental states we refer to when we say things like “Geri believes that abortion is wrong” are beliefs and those who claim that those states are desires, emotions,

³⁵ Ibid., “Expressivism and Internalism,” 343.

³⁶ Recall that the first objection can also be read in this way.

or pro/con attitudes. So if the first premise of the above argument is correct, it is possible, at least in principle, to settle the debate between mental state cognitivists and mental state non-cognitivists empirically in the way Joyce describes: “take some persons who are paradigmatic instances of ‘moral judges’ and with a PET scan watch for neural evidence of a ‘conative state’—say, activity in the amygdala (or whatever)—when we prompt them to think about euthanasia, or Adolph Hitler, or returning the Elgin marbles.” Now by “conative state” Joyce presumably means “motivation implicating state.” But we saw above that *both* mental state cognitivists and mental state non-cognitivists can hold that moral judgments (construed as mental states) are motivation implicating states. So the mental state cognitivism/mental state non-cognitivism debate cannot be settled empirically in the way that Joyce describes. The first premise of the above argument is, therefore, false. It is not the case that the cognitivism/non-cognitivism debate can be settled empirically in the way that Joyce describes if the debate is a debate about the nature of certain mental states.

Joyce might respond by claiming that just as it is in principle possible to use PET scans to determine whether a given mental state is a motivation implicating state, it is in principle possible to use PET scans to determine whether a given mental state is a belief. He could then restate his objection as follows. If mental state non-cognitivism is the view that moral judgments (construed as mental states) are motivation implicating states *and* are not beliefs, then it is in principle possible to settle the mental state cognitivism/mental state non-cognitivism debate using PET scans. So it is possible, at least in principle, to settle the cognitivism/non-cognitivism debate empirically using PET scans if that debate is a debate about the nature of certain mental states. But the cognitivism/non-cognitivism

debate cannot, even in principle, be settled empirically using PET scans. So the cognitivism/non-cognitivism debate is not a debate about the nature of certain mental states.

Now it is not obvious that it is possible, even in principle, to determine using PET scans whether a given mental state is a belief. But let's grant for the sake of argument that it is. Let's also assume that it is in principle possible to determine whether a given mental state is a motivation implicating state using PET scans. Finally, let's assume that the *only* theses essential to mental state non-cognitivism are that moral judgments (construed as mental states) are motivation implicating states and that they are not beliefs. Given these assumptions, it is indeed possible, at least in principle, to settle the mental state cognitivism/mental state non-cognitivism debate empirically using PET scans. Joyce's worry is that it is implausible that *the cognitivism/non-cognitivism debate* can be settled in this way. If that is implausible, then Joyce is right that the issue between cognitivists and non-cognitivists cannot be identified with a debate about the nature of certain mental states, that is, with the mental state cognitivist/mental state non-cognitivist debate.

But *is it* implausible that the expressivist/cognitivist debate can, at least in principle, be settled empirically using PET scans? The proper response to this question, I think, is to deny that there is really such a thing as *the* cognitivist/non-cognitivist debate. Some philosophers have as their primary concern the nature of the mental states that we refer to when we say things like "Geri believes that abortion are wrong." One can be either a cognitivist or a non-cognitivist about these states. In this context, cognitivism is mental state cognitivism and non-cognitivism is mental state non-cognitivism. We are

now assuming, if only for the sake of argument, that *this* debate *can* in principle be settled empirically using PET scans. Other philosophers are primarily concerned with moral utterances. Here too one can be either a cognitivist or a non-cognitivist. But in this context cognitivism is linguistic cognitivism and non-cognitivism is linguistic non-cognitivism. Let us assume for the sake of argument that *this* debate *cannot* be settled empirically using PET scans. So there are at least at two cognitivism/non-cognitivism debates—a fact that is obscured by the ambiguity of “moral judgment.” One of these debates concerns what certain *mental states are* and can—we are assuming—be settled empirically using PET scans, at least in principle. The other concerns what certain *utterances express* and cannot—we are assuming—be settled empirically using PET scans, even in principle. Is there any reason to think that one of these debates is *the real* expressivism/cognitivism debate? If there is, I don’t see it.

Conclusion

In this chapter I distinguished linguistic non-cognitivism and mental state non-cognitivism. I explained why this distinction matters. Linguistic non-cognitivists have a reply to the objection from amoralism that mental state non-cognitivists do not have. Furthermore, mental state non-cognitivism sidesteps the potentially problematic issue of giving an account of the “express” relation. These two considerations are enough to indicate that philosophers should be clear about whether they have in mind mental state non-cognitivism, linguistic non-cognitivism, or both. I then presented three arguments given by Richard Joyce against construing non-cognitivism as mental state non-cognitivism and showed that each argument fails. For all Joyce’s arguments show, mental state non-cognitivists can appeal to motivation internalism to support their view, mental

state non-cognitivism does not imply that a moral judgment was not made in cases where intuitively one has been made, and construing the cognitivism/non-cognitivism debate mentalistically does not render that debate empirical in a problematic or “unwelcome” way.

Chapter Three

Can Internalism Explain the Practical Character of Morality?

Introduction

This chapter is about judgment internalism, the view that there is a conceptual connection between moral judgment and motivation.³⁷ The strongest version of this view holds that it is conceptually impossible for someone to have a moral belief and lack a corresponding overriding motivation.³⁸ Such a strong internalism implies, for example, that the claim that Jones believes that stealing is wrong but is not averse to stealing is on par with the claim that Jones is a married bachelor. Indeed, it implies that there is a conceptual guarantee that Jones' motivation not to steal is stronger than any conflicting motivation he may have. This very strong form of internalism is widely regarded as implausible even by internalists, and contemporary discussions usually focus on weaker

³⁷ More precisely, judgment internalism is the view that there is a *non-trivial* conceptual connection between moral judgment and motivation. See Hallvard Lillehammer, "Smith on Moral Fetishism," *Analysis* 57 (July 1997):187. Judgment internalism is sometimes identified with the view that there is a necessary connection between moral judgment and motivation. See, for example, Sigrún Svavarsdóttir, "Moral Cognitivism and Motivation," *The Philosophical Review* 108 (April 1999):162. Since a necessary connection need not be a *conceptual* connection, this differs from my characterization. However, my characterization is also common in the literature. For example, Hallvard Lillehammer writes: "According to internalism about moral judgments there is an interesting conceptual connection between an agent's making a moral judgment and that agent's motivation. The externalist denies this and claims that any interesting connection between moral judgment and motivation is contingent." Hallvard, "Smith on Moral Fetishism," 187. In any event, I am here specifically concerned with the view that moral judgments come with motivation as a matter of *conceptual* necessity. So there are two ways for judgment internalism as I am understanding it to turn out false: it is false if moral judgment is only contingently connected to motivation, and it is false if the two are necessarily connected but the modality is something other than conceptual necessity. For example, it is false if the connection between moral judgment and motivation is one of metaphysical necessity.

³⁸ I am using "belief" in a broad sense that is neutral between cognitivism and non-cognitivism, neutral, that is, on the question of whether moral judgments *really are* beliefs. I am using 'moral judgment' and 'moral belief' interchangeably.

versions.³⁹ I do two things here. First, I argue that while two popular weaker versions of internalism sidestep common objections to the view, they are *too weak* to explain at least one aspect of the practical character of morality, something that is often thought to provide internalism with *prima facie* support. Second, I explore the question of whether there is some other weaker version of internalism that avoids the central criticism of the strongest version while remaining strong enough to be able to explain that aspect of the practical character of morality. I identify different versions of internalism that fit this bill. As my goal for this part of the chapter is to simply delineate some options for internalists who hope to derive support from the practical character of morality, I leave open the question of whether any of these versions of internalism is true, as well as the comparative question of which version is most plausible.

The Practical Character of Morality

In *Moral Realism and the Foundations of Ethics*, David Brink describes what he calls “the practical character of morality”:

Moral considerations are practical in some very important sense. Agents engage in moral deliberation in order to decide what to do and give moral advice with the aim of influencing others’ conduct in certain ways. We expect people who accept moral claims or make moral judgments to act in certain ways. We would regard it as odd for people who accepted moral claims about an issue to be completely indifferent about that issue. For these reasons, we expect moral considerations to motivate people in certain ways, or at least to provide them with reason to act in those ways.⁴⁰

³⁹ For example, Russ Shafer-Landau dismisses this type of internalism as “obviously false.” Shafer-Landau, *Moral Realism: A Defence*, 142. Shafer-Landau is an externalist. Internalists who reject this strong version of internalism include Michael Smith, James Dreier, and Mark van Roojen. Smith, *The Moral Problem*, 61; James Dreier, “Internalism and Speaker Relativism,” *Ethics* 101 (October 1990): 9-10; Mark van Roojen, “Humeanism and Anti-Humeanism About Moral Judgments,” *Philosophy and Phenomenological Research* 65 (July 2002): 34-35.

⁴⁰ Brink, *Moral Realism and the Foundations of Ethics*, 37.

It is important to see that the idea that morality is practical in the way Brink here describes is consistent with the falsity of internalism. Internalism holds that there is a *conceptual connection* between moral judgment and motivation, but the phenomena that Brink cites may all have their source in *contingent* facts about human psychology. Asserting that morality has a practical character does not, therefore, beg the question against judgment externalism, the view that there is only a contingent connection between having a moral belief and having a corresponding motivation.⁴¹ And there is, in fact, broad agreement among *both* internalists and externalists that morality has a practical character. (Brink himself is an externalist.) The dispute between internalists and their externalist opponents is thus not over whether moral judgment and motivation are intimately connected. Rather, the dispute is over whether the intimate connection between moral judgment and motivation that both parties acknowledge is grounded in a conceptual truth.

Although the practical character of morality does not imply internalism, it is sometimes suggested that internalism derives *prima facie* support from it.⁴² This suggestion rests on two ideas. The first is that the practical character of morality is a datum that any adequate theory of moral judgment will accommodate. Brink, for example, clearly accepts this idea. He presents his above remarks as commonsense observations about moral practice and is eager to show that they fit well with his own theory.⁴³ This idea is also evident in Russ Shafer-Landau's discussion of the

⁴¹ Notice that judgment internalism as I am understanding it and the view that I have just identified as judgment externalism can both be false. This would be the case if moral judgment and motivation were necessarily connected but the modality was something other than conceptual necessity. So on my way of carving things up, judgment externalism turns out not to be the denial of judgment internalism. This is an unfortunate result, but I do not think that anything substantive hangs on it.

⁴² See, for example, Shafer-Landau, *Moral Realism: A Defence*, 156.

⁴³ Brink, *Moral Realism and the Foundations of Ethics*, 37.

internalism/externalism debate in his *Moral Realism: A Defence*. Shafer-Landau treats the practical character of morality as part of the appearances, a feature of moral phenomenology, and tells us that if “externalists cannot account for the obviously close connection between the making of a moral judgment and the motivation to adhere to it, then they are in trouble.”⁴⁴

The second idea is that internalists seem to be better placed than externalists when it comes to explaining this practical character of morality. They can point to the conceptual connection they claim obtains between moral judgment and motivation: Morality is practical in the ways Brink describes because it is “built into” the concept of a moral belief that a person counts as having one only if that person has a corresponding motivation. In contrast, it is not as clear why morality has a practical character if it is a wholly contingent matter whether someone who has a moral belief has any motivation to act as the belief prescribes. Shafer-Landau sums up the point: “Here, appeal to the phenomenology of moral experience seems to support internalism rather than externalism. Moral judgments do exert a pull over us, and internalists have a simple, straightforward explanation of this. The externalist story cannot be so simple.”⁴⁵

The point I want to make is that this thought that internalism provides a “simple, straightforward explanation” of the practical character of morality does not apply to two popular weaker forms of the view. More precisely, I will show that these weaker versions of internalism cannot explain at least one of the central ways in which morality is practical. As Brink observes, we “expect people who accept moral claims or make moral

⁴⁴ Shafer-Landau, *Moral Realism: A Defence*, 156; see also van Roojen, “Humean and Anti-Humean Internalism About Moral Judgments,” 32.

⁴⁵ Shafer-Landau, *Moral Realism: A Defence*, 156; see also “Humean and Anti-Humean Internalism About Moral Judgments,” 32.

judgments to act in certain ways.”⁴⁶ If I learn, for example, that you believe that eating meat is immoral, I will expect you to avoid eating meat. Shafer-Landau identifies a closely related phenomenon when he writes: “we suspect the sincerity of someone who proclaims fidelity to a moral code, all the while showing no inclination to abide by it.”⁴⁷ To stick with the example, if you tell me that you think it is immoral to eat meat but show no inclination to avoid eating meat, it would be natural for me to wonder whether you are putting me on. This suspicion that Shafer-Landau identifies is presumably a consequence of still another phenomenon: we have a general reluctance to attribute a moral belief to someone who shows no sign of being motivated to act as the belief prescribes.⁴⁸ That is the practical aspect of morality I will focus on. Weak versions of internalism are popular because they sidestep putative counterexamples that are widely held to refute stronger versions. Yet, in sidestepping the putative counterexamples they also lose any ability to explain why we resist attributing a moral belief to someone who seems to lack any corresponding motivation. Or so I will argue.

Two Weaker Versions of Internalism

Recall that the strongest version of internalism holds that it is conceptually impossible for someone to have a moral belief and fail to have a corresponding overriding motivation, that is, a corresponding motivation that is stronger than any conflicting motivations the person may have. Now *this* version of internalism *can* explain our general reluctance to attribute a moral belief to someone who shows no sign of being motivated to act as the belief prescribes. For if it is conceptually impossible to have a

⁴⁶ Brink, *Moral Realism and the Foundations of Ethics*, 37.

⁴⁷ Shafer-Landau, *Moral Realism: A Defence*, 156.

⁴⁸ A virtue of focusing on this datum as opposed to the one that Shafer-Landau identifies is that it leaves open the possibility that one can sincerely express a mental state that one is not in. See Joyce, “Expressivism and Internalism,” 336-344. See also Ridge, “Sincerity and Expressivism.”

moral belief without having a corresponding overriding motivation, then just as we will resist believing that Jones is a bachelor if we think that Jones is married, we will resist attributing a moral belief to someone we think does not have a corresponding overriding motivation. And since we know that overriding motivations tend to reveal themselves in behavior, if someone shows no sign of being motivated to act in a certain way, we will think that that person does not have an overriding motivation to act in that way.

However, as I mentioned earlier, this very strong version of internalism is widely dismissed as implausible. Everyone acknowledges that people often have motivations that conflict with their moral beliefs. The question is whether it is conceptually possible for a conflicting motivation to be stronger than the motivation to act as the moral belief prescribes. The strongest version of internalism tells us that that is not conceptually possible. Yet, it seems obvious that such cases *are* conceptually possible. Indeed, many would say that such cases actually occur all of the time.⁴⁹ So while the strongest form of internalism can explain our datum, there is a general consensus that it is false.

I turn now to weaker versions of internalism. One weaker version that receives a lot of attention is identical to the strongest version with one exception: it holds that when someone has a moral belief what is conceptually guaranteed is not that the person has a corresponding *overriding* motivation but just that the person has *some* motivation to act as the belief prescribes.⁵⁰ Since it concedes the conceptual possibility of cases in which a person's motivation to act as one of her moral beliefs prescribe is weaker than other, conflicting motivations the person has, this version of internalism avoids the central criticism of the strongest version. But can it explain our general reluctance to attribute a

⁴⁹ Dreier, for example, seems to think this. Dreier, "Internalism," 10.

⁵⁰ Shafer-Landau devotes an entire chapter to this version of internalism. Shafer-Landau, *Moral Realism: A Defence*, 142-161.

moral belief to someone who shows no sign of having a corresponding motivation? I will now argue that it cannot.

The problem is that this version of internalism is consistent with someone's having a moral belief and having only an extremely weak corresponding motivation. The only thing it rules out as conceptually impossible is someone's having *no motivation whatsoever* to act as one of one's moral belief prescribes.⁵¹ Now, the alleged conceptual impossibility of such cases can explain why we resist attributing a moral belief to someone we think has no motivation to act in accordance with the belief. But notice that that is not quite what we wanted explained.⁵² The question was not whether this weaker version of internalism can explain our general reluctance to attribute a moral belief to someone we think *has no* corresponding motivation. Rather, the question was whether it can explain our reluctance to attribute a moral belief to someone who *shows no sign of* having a corresponding motivation. The version of internalism under consideration explains *that* only if it is the case that we typically infer from someone's showing no sign of being motivated to act in a certain way that the person has no motivation to act in that way. But we do not typically draw that conclusion, and for good reason: we know that someone's failure to show any sign of being motivated to act in a certain way may be due *not* to the *absence* of any motivation to act in that way but rather to the *weakness* of the motivation relative to the person's other motivations.

⁵¹ See Dreier, "Internalism," 12.

⁵² Of course, that is not to say that we are not reluctant to attribute a moral belief to someone who has no corresponding motivation, or that that is not also a datum that needs to be explained. I will return to this point.

Another popular version of internalism starts from the thought that the version we have just been considering is still too strong.⁵³ As we noted, that version of internalism is weaker than the strongest version in that it grants the conceptual possibility of someone's having a moral belief while lacking a corresponding overriding motivation. What it denies, recall, is that it is conceptually possible for someone to have a moral belief and have no motivation whatsoever to act as the belief prescribes. But are such cases really conceptually impossible? Many theorists think not. Many theorists think, for example, that states such as severe depression and physical or emotional exhaustion can leave a person entirely unmotivated to do something that she genuinely believes she is morally required to do.⁵⁴ There is a worry, then, that even a version of internalism that requires only that people have *some* (possibly very weak) motivation to act in accordance with their moral beliefs is subject to counterexample. Some internalists have responded to this worry by arguing that the conceptual connection between moral judgment and motivation is *defeasible*. According to these theorists, it is a conceptual truth that *in normal cases* a person who has a moral belief has some (possibly very weak) corresponding motivation. In slightly different terms, the idea is that it is conceptually impossible *other things being equal* for someone to have a moral belief and have no motivation whatsoever to act as the belief prescribes.⁵⁵

Although this thesis claims that moral judgment is only defeasibly connected to motivation, it gives the alleged defeasible connection the status of a conceptual truth. So it counts as genuine form of internalism. Yet, because it acknowledges the conceptual

⁵³ This view is held, for example, by Smith, Dreier, and van Roojen. Smith, *The Moral Problem*, 61; Dreier, "Internalism," 9-10; van Roojen, "Humean," 34-35.

⁵⁴ See Smith, *The Moral Problem*, 120-121 and Dreier, "Internalism," 10.

⁵⁵ Smith, Dreier, and van Roojen are again prominent examples. Smith, *The Moral Problem*, 61-63; Dreier, "Internalism," 14; van Roojen, "Humean," 35-36.

possibility of moral belief without any corresponding motivation, it is harder to refute than the previous version of internalism. For example, it is not refuted by the (putative) conceptual possibility of cases in which severe depression or exhaustion leaves a person with entirely no motivation to act in accordance with one of her moral beliefs. The defender of this version of internalism need only claim that cases that involve such severe depression or exhaustion are “abnormal,” that they are cases in which other things are not equal. However, as I will now argue, this version of internalism *is* like the previous version in being unable to explain why we are generally reluctant to attribute a moral belief to someone who shows no sign of having a corresponding motivation.

Let me begin by noting a complication. The complication stems from the fact that this version of internalism can itself take many different forms depending on what one says about the defeasibility conditions. Some theorists cash out the defeasibility conditions in explicitly normative terms. Michael Smith, for example, famously argues that it is conceptually necessary that someone who judges an act to be right will have some motivation to perform that act *unless she is practically irrational*.⁵⁶ Other theorists consider cashing out the defeasibility condition in non-normative terms. For example, James Dreier considers, but ultimately rejects, taking a statistical approach.⁵⁷ The idea is that it is conceptually necessary that moral belief *usually* comes with some corresponding motivation.⁵⁸ Another approach that Dreier considers but rejects is the brute list

⁵⁶ Smith, *The Moral Problem*, 61.

⁵⁷ Dreier, “Internalism,” 7-8.

⁵⁸ One might have questions about this kind of conceptual statistical truth. If it is a conceptual truth that most F’s are G’s, on what basis will we decide whether some particular F is one of the rare non-G F’s? As Brian Kierland has pointed out to me, we would at the very least need some way of grouping F’s that does not rely on determining whether something is a G. I assume here that there are good answers to these and other questions one might raise about the possibility of this sort of truth.

approach.⁵⁹ Someone who takes this approach tries to simply list all of the considerations that can defeat the connection between moral judgment and motivation. Dreier himself takes a still different tack. He tries to use hypothetical cases to show that we have a grip on a conception of a normal case that renders the version of internalism under consideration plausible while admitting that he cannot give an adequate analysis of that conception in either normative or non-normative terms.⁶⁰

Now, given that the version of internalism under consideration can take such different forms, it might seem like a mistake to ask in general terms whether it can explain our datum. One might think that we should instead ask whether a normative form of this version of internalism can explain the datum, whether a statistical form can explain it, and so on. It turns out, however, that while it may be an interesting issue what defenders of this version of internalism should say about the defeasibility conditions, that issue does not matter for our immediate purpose. For we are already in a position to see that *however* that issue is resolved, the version of internalism under consideration cannot explain why we resist attributing a moral belief to someone who shows no sign of having a corresponding motivation. The crucial point is that—*whatever* one says about the defeasibility conditions—in cases in which the alleged conceptual connection between moral judgment and motivation is *not* defeated the only thing that is conceptually guaranteed is that the person has *some* (possibly very weak) motivation. But we saw when discussing the previous version of internalism that a conceptual guarantee of only *some* (possibly very weak) motivation cannot explain our reluctance to attribute a moral belief to someone who shows no sign of having a corresponding motivation.

⁵⁹ Dreier, “Internalism,” 12-13.

⁶⁰ Dreier, “Internalism,” 13-14. Simon Blackburn takes what appears to be the same approach. Blackburn, *Ruling Passions*, 59-68.

We have just seen that two popular weaker versions of internalism cannot explain why we are generally reluctant to attribute a moral belief to someone who shows no sign of having a corresponding motivation. In both cases, the problem is that the alleged conceptual connection is between moral judgment and only some, possibly very weak, motivation. This shows that internalists who want to claim support from their ability to explain our reluctance to attribute moral belief in the absence of any sign of corresponding motivation need to hold that a person's having a moral belief is conceptually linked to more than just that person's having *some* motivation to act as the belief prescribes. Of course, if the internalist's *only* concern were to explain our reluctance to attribute moral belief where we do not see any sign of motivation, then she could simply embrace the strongest version of internalism, the view that it is conceptually impossible to have a moral belief and lack a corresponding *overriding* motivation. For we saw that that version of internalism can explain the datum. Yet, as I noted, that version of internalism is widely regarded as *too strong*. So a question naturally arises: Is there a version of internalism that allows for the conceptual possibility of moral belief without overriding motivation that is also strong enough to be able to explain our reluctance to attribute a moral belief to someone who shows no sign of being motivated to act as the belief prescribes? The remainder of this chapter will focus on that question. But before I turn to that question I want to first briefly address an objection that can be raised against the preceding argument that the two weaker versions of internalism we have already discussed cannot explain our datum.

I have argued that the two popular weaker versions of internalism just discussed cannot explain our reluctance to attribute a moral belief to someone who shows no sign

of having a corresponding motivation. The objection I have in mind concedes that point but questions its significance for the broader issue of whether weak versions of internalism can explain the practical character of morality. The problem, according to the objection, is that it is a mistake to identify the practical character of morality with our reluctance to attribute a moral judgment to someone who shows no sign of having a corresponding motivation. Furthermore, the objection continues, what the practical character of morality *does* consist in is something that one of the weaker versions of internalism *can* explain. Exactly how this objection goes will depend, of course, on what the objector identifies as the practical character of morality. But it is easy to think of accounts of the practical character of morality that the weaker versions of internalism can explain. For example, suppose we say that the practical character of morality consists in our having a general reluctance to attribute a moral belief to someone we think has no corresponding motivation. (Recall the above point that there is a difference between *having no motivation* to act in a certain way and *showing no sign* of being motivated in that way.) We have already noted that the first of the two weaker versions of internalism can help explain *that*.⁶¹ Or suppose that the practical character of morality consists in our being reluctant to attribute a moral belief to someone we think has no corresponding motivation *and* who we think is not severely depressed or exhausted.⁶² In that case, the second weaker version of internalism can explain the practical character of morality.

This objection raises an obvious but important point: whether a particular version of internalism can explain the practical character of morality depends on what we identify as the practical character of morality. A consequence of this point is that a proponent of a

⁶¹ See the first full paragraph on p. 31.

⁶² See Dreier, "Humean," 32.

particular version of internalism can reply to *any* attempt to show that her version of internalism cannot explain the practical character of morality by claiming that the practical character of morality has been incorrectly described. How, then, can we proceed? I see two broad options. First, we might argue that some characterization of the practical character of morality is, in fact, the correct characterization. If it then turns out that a particular version of internalism cannot explain the identified datum, then so much the worse for that version of internalism: That version does not derive support from an ability to easily explain the practical character of morality. For example, we could argue that the practical character of morality consists in the datum we have been considering, namely, our reluctance to attribute a moral belief to someone who shows no sign of having a corresponding motivation. If we take this route, then we will respond to the objection by simply denying the claim that I have mischaracterized the practical character of morality.

A problem with the first option is that morality may be practical in more than one way. It may be that the expression “the practical character of morality” is best thought of as a general heading under which multiple related phenomena fall. This suggests a different approach: we can try to identify at least *one of the ways* in which morality is practical and then ask of various versions of internalism whether they can explain *that* datum. That is the approach I have taken here. I have not tried to show that the two weaker versions of internalism cannot explain the practical character of morality. Rather, I have argued for the more modest conclusion that they cannot explain *at least one aspect* of the practical character of morality. So for the objection to engage my argument it would have to either deny that we are generally reluctant to attribute a moral belief to

someone who shows no sign of having a corresponding motivation or concede that we are so reluctant but deny that this reluctance is an aspect of the practical character of morality. Yet, it seems to me that we do have such a general reluctance, and I do not see how it could fail to count as an important aspect of the practical character of morality.

Very Strong Internalism and Overriding Defeasible Internalism

I turn now to the question of whether there is a version of internalism that allows for the conceptual possibility of moral belief without overriding motivation that is also strong enough to be able to explain our reluctance to attribute a moral belief to someone who shows no sign of being motivated to act as the belief prescribes. The answer is that there are multiple versions of internalism that fit this bill. One such version is a version of internalism that to my knowledge has not yet been considered in the literature. I call it “very strong indefeasible internalism.” As the name suggests, the idea is that there is an indefeasible conceptual connection between moral judgment and having a very strong motivation to act as the belief prescribes. Specifically, this version of internalism claims that it is a conceptual truth that someone who has a moral belief has a very strong corresponding motivation.⁶³ Now, since someone can have a very strong motivation to act in a certain way without having an *overriding* motivation to act in that way, this version of internalism allows for the conceptual possibility of someone’s having a moral belief without having an overriding motivation to act as the belief prescribes. This version of internalism thus avoids the central criticism of the strongest version. Yet, it is

⁶³ When does a motivation count as “very strong”? I think we have an intuitive grip on this idea. However, one point is worth clarifying: we can understand “very strong” in either comparative or absolute terms. Understood comparatively, a very strong motivation is one of the stronger motivations some person has. Understood absolutely, a very strong motivation meets some (vague) threshold of strength. A motivation can be very strong in the first sense but not the second, and vice versa. I have in mind the absolute sense. Thanks to Brian Kierland for asking me to be clearer on this point.

like the strongest version in being able to explain our reluctance to attribute moral belief where we do not see any sign of corresponding motivation. For if it is conceptually impossible to have a moral belief without having a very strong motivation to act as the belief prescribes, then we will resist attributing a moral belief to someone we think does not have a very strong corresponding motivation. And while we do not typically infer that someone has *no* motivation to act in a certain way from the fact that the person shows no sign of being motivated to act in that way, we *do* typically infer from someone's showing no sign of being motivated to act in a certain way that that person does not have a *very strong* motivation to act in that way.

Although very strong indefeasible internalism avoids the central criticism of the strongest version of internalism, many theorists will still reject it as too strong. For we saw that even some internalists think that it is conceptually possible for someone to have no motivation whatsoever to act in accordance with one of her moral beliefs. And if it is conceptually possible for someone to have a moral belief and lack *any* corresponding motivation, then it is, of course, conceptually possible for someone to have a moral belief and lack a *very strong* corresponding motivation. So we should ask whether there is a version of internalism weaker than very strong indefeasible internalism that can explain our datum.

An obvious suggestion is a defeasible version of very strong internalism. Here the idea is that it is a conceptual truth that *in normal cases* someone who has a moral belief will have a very strong motivation to act as the belief prescribes. The suggestion, in other words, is that *other things being equal* it is conceptually impossible for someone to have a moral belief and lack a very strong corresponding motivation. Because this version of

internalism is consistent with the conceptual possibility of cases in which someone has a moral belief without having a very strong motivation to act as the belief prescribes it avoids the objection that many will have to very strong indefeasible internalism. But can it explain why we are generally reluctant to attribute a moral belief to someone who shows no sign of having a corresponding motivation?

Unlike with the defeasible version of internalism discussed in the previous section—the version which claims that other things being equal it is conceptually impossible to have a moral belief and have no motivation whatsoever to act as the belief prescribes—here it does matter what one says about the defeasibility conditions. On at least one approach to spelling out what counts as a “normal case,” a case where other things are equal, very strong defeasible internalism can explain our datum. The approach I have in mind is the statistical approach. When “normal case” is cashed out in statistical terms, very strong defeasible internalism becomes the claim that it is a conceptual truth that moral belief *usually* comes with a very strong corresponding motivation.⁶⁴ Suppose this claim is true. Presumably, if it is a conceptual truth that moral belief usually comes with a very strong corresponding motivation, then it is also a conceptual truth that we are unlikely to come across a case in which someone has a moral belief but does not have a very strong motivation to act as the belief prescribes. And if that is a conceptual truth, then we will resist attributing a moral belief to someone we think does not have a very strong corresponding motivation. When we then add the point that we typically infer from someone’s showing no sign of being motivated to act in a certain way that that person does not have a very strong motivation to act in that way it becomes easy to see

⁶⁴ See footnote 58.

why we are reluctant to attribute a moral belief to someone who shows no sign of having a corresponding motivation.

So a statistical form of very strong defeasible internalism can explain our datum. Matters are not so clear, however, with other forms of this version of internalism. Suppose, for example, that we cash out the defeasibility conditions in explicitly normative terms. More specifically, suppose that we follow Smith and say that the conceptual connection between moral judgment and motivation can be defeated by practical irrationality. This turns very strong defeasible internalism into the following thesis: it is a conceptual truth that a person who has a moral belief has a very strong motivation to act as the belief prescribes *unless she is practically irrational*. As I will now explain, whether this thesis can explain our reluctance to attribute a moral belief to someone who shows no sign of having a corresponding motivation depends on whether we think cases of practical irrationality are rare.

We have already noted that when someone shows no motivation to act in a certain way we typically infer that the person lacks a very strong motivation to act in that way. If the version of internalism under consideration is true, we will naturally move from this thought that the person lacks a very strong motivation to act in the relevant way to the disjunctive thought that the person *either* lacks a corresponding moral belief *or* has a corresponding moral belief but is practically irrational. Now suppose that we think cases of practical irrationality are rare. More precisely, suppose that we are generally averse to concluding that someone is practically irrational. In that case, we will be inclined to go on to infer that the person lacks a corresponding moral belief. So if we are, in fact, generally averse to believing that a person is practically irrational, this normative form of

very strong defeasible internalism can help explain why we are reluctant to attribute a moral belief to someone who shows no sign of having a corresponding motivation. But now suppose that we think cases of practical irrationality are common and are thus often quite happy to believe that a person is practically irrational. Given this assumption, when someone shows no sign of having a motivation to act in a certain way we may be just as likely to conclude that the person has a corresponding moral belief but is practically irrational as we are to conclude that the person lacks a corresponding moral belief. Our reluctance to attribute a moral belief to such a person would, therefore, remain unexplained.

Consider now the brute list approach to the defeasibility conditions. Recall that internalists who take this approach attempt to simply list the considerations that can defeat the (alleged) conceptual connection between moral judgment and motivation. Applied to very strong defeasible internalism this approach gives us the claim that it is a conceptual truth that someone who has a moral belief has a very strong corresponding motivation *unless ...*, where the ellipsis is filled in with a list of the considerations one takes to be potential defeaters. For example, the ellipsis might be filled in with “the person is severely depressed or is emotionally or physically exhausted.” Can a very strong defeasible internalism that takes this form explain why we are reluctant to attribute a moral belief to someone who shows no sign of having a corresponding motivation?

The answer is that it depends on what is included on the list of potential defeaters. In particular, it depends on whether the list includes any states that we take to be common. Suppose that it does contain such a state. Call this state “S.” In that case, this form of very strong defeasible internalism fails to explain our datum for the same reason

that the normative form fails to explain it if we do not think it is rare for someone to be practically irrational: when we encounter someone who shows no sign of being motivated to act in a certain way we may be just as likely to conclude that that person has a corresponding moral belief but is in S as we are to conclude that the person does not have a corresponding moral belief. On the other hand, if the list of potential defeaters only includes states that we think are rare, then this form of very strong defeasible internalism explains our datum for the same reason that the normative form explains it on the assumption that it is rare for someone to be practically irrational. For if the version of internalism under consideration is true, then when we encounter someone who shows no sign of being motivated to act in a certain way we will naturally think that this person *either* lacks a corresponding moral belief *or* has a corresponding moral belief but is in one of the potentially defeating states. And if we think that it is rare for someone to be in any one of the potentially defeating states, we will then be inclined to infer that this person simply lacks a corresponding moral belief.

I want to consider one final approach to the defeasibility conditions. I mentioned earlier that Dreier argues that there is a sense of “normal case” that renders a defeasible version of internalism plausible while admitting that he cannot give an adequate analysis of that conception. Perhaps the best way to understand this approach to the defeasibility conditions is to look at Dreier’s argument. Dreier begins by presenting two putative counterexamples to indefeasible versions of internalism. The first comes from Michael Stocker’s. Dreier writes:

... Stocker describes a politician who cared in his youth ‘about the suffering of people in all parts of the world and devoted himself to making their lives better. But now he concerns himself only with the lives and fortunes of his close family and friends. He remembers his past, and he

knows that there is still a lot he could do to help others. But he no longer has any desire to do so.' Of course, there are different ways of explaining the politician's situation. Maybe he has ceased to believe that it is a good thing to help others But Stocker's insight is that it makes perfect sense to describe the uncaring man another way. Couldn't he say, 'I have not changed my moral beliefs; I still think the morally good thing to do is to help the unfortunate. But I am no longer interested in doing what is morally good.'⁶⁵

The second putative counterexample revolves around a group of people, as opposed to a single individual. Dreier presents it as follows:

The Sadists are a group of people who recognize what sorts of actions, states of affairs, characters, their society calls 'morally good.' They recognize them, but they hate them. They find those things repulsive and avoid them at all costs. In general, they are not in the slightest motivated to perform those actions. Here is the way they put it: 'We despise what is good because it is good. We believe the things you say are good are in fact, and we are never motivated to promote those things. Quite the opposite!' That is what the Sadists say, and it seems as though it might be true.⁶⁶

Dreier thinks that these two cases constitute genuine counterexamples to the idea that it is conceptually impossible to have a moral belief and lack a corresponding motivation. This leads him to reject indefeasible internalism in favor of the second of the weaker versions of internalism discussed in the previous section, the view that it is a conceptual truth that in normal cases someone who has a moral belief has some corresponding motivation.

Dreier thinks that a key to defending this version of internalism is showing that "we have some independent conception of normality, a conception of the normal that does not depend on a prior notion of 'morally good.'"⁶⁷ To show that we do have such a conception Dreier contrasts the original Stocker and Sadist cases with modified versions.

Here is Dreier's argument in his own words:

⁶⁵ Dreier, "Internalism," 6. The quotation from Stocker is from Michael Stocker, "Desiring the Bad: An Essay in Moral Psychology," *The Journal of Philosophy* 76 (December 1979): 741.

⁶⁶ Dreier, "Internalism," 10-11.

⁶⁷ *Ibid.*, 13.

Suppose the politician to have been born to pioneers, and to have spent his life in the company of rugged individualists. And suppose he now says, 'What my friends believe is wrong: not individualism but a life in the service of others is really good.' But the politician has no inclination to serve the less fortunate; instead he advances the cause of self-reliance whenever he can.

Or suppose we discover an isolated culture of mean-spirited folk who go out of their way to cause harm and humiliation at every opportunity, while shunning kindness and fairness and generally avoiding any action that would contribute to the well-being of sentient creatures. Unlike the Sadists in our earlier example, they have never met people who feel differently. These Sadists call the things that attract them 'gad' and those things that repel them 'bood.'

It seems to me that what we want to say about the new politician is that he is using the word 'good' either insincerely or incorrectly. We will not take his assertion at face value and attribute to him the belief 'Life in the service of others is good.' Similarly, what we want to say about the isolated Sadists is that their 'gad' means good and 'bood' means bad. And if I am right about our intuitions in these new cases, then a satisfactory account of the difference between them and the original examples is that in each of the originals there was a salient standard of normality against which to contrast the peculiar, whereas in the modified examples there were no such standards. So we can distinguish among cases in which agents are not motivated to do what we think is good, recognizing on the one hand genuine exceptions to ... [indefeasible] internalism and on the other insincerity or failure properly to understand or use moral terms. Since we can make that distinction, our conception of normality is not merely, 'Those circumstances under which a person is motivated by what she believes to be good.' We must have some independent idea of normality.⁶⁸

After presenting this argument Dreier immediately goes on to concede that he does "not know how to specify this independent conception rigorously."⁶⁹ He writes:

It is clear to me that if everyone in a community behaves a certain way, then that behavior is normal for the community, and if a person has a certain state of character for all of her life, then behavior flowing from that state is normal for her. But that is only to specify a limit. Though I think I have successfully argued that we do have a grip on the conception I need, I cannot now provide an analysis.⁷⁰

⁶⁸ Ibid.

⁶⁹ Ibid., 14.

⁷⁰ Ibid.

We can set aside the interesting question of whether Dreier's argument works. For our purposes the important point is that one can defend a defeasible version of internalism without attempting to articulate the conditions under which the alleged conceptual connection between moral judgment and motivation can be defeated. Dreier's argument nicely illustrates how to do this: one need only maintain that whenever the (alleged) conceptual connection breaks down one is dealing with a case that is abnormal according to *some* "salient standard of normality." Let us now apply this approach to defeasibility conditions to very strong defeasible internalism and see whether it yields a form of internalism that can explain our general reluctance to attribute a moral belief to someone who shows no sign of having a corresponding motivation.

The version of internalism we need now to consider amounts to the claim that it is a conceptual truth that someone who has a moral belief has a very strong motivation to act as the belief prescribes *unless the case is abnormal according to some salient standard of normality*. Now, just as the normative form of very strong defeasible internalism can explain our datum if, but only if, we think that cases of practical irrationality are rare, and just as a brute list form can explain it if, but only if, the list of potential defeaters only includes states that we think are rare, so this form of very strong defeasible internalism can explain our reluctance to attribute a moral belief to someone who shows no sign of having a corresponding motivation if, but only if, we think that cases that are abnormal according to some salient standard of normality are rare. For if the version of internalism under consideration is true but we think that cases that are abnormal according to some salient standard of normality are common, then when we encounter someone who shows no sign of being motivated to act in a certain way we may

be just as likely to conclude that the person has a corresponding moral belief but is abnormal according to some salient standard of normality than we are to conclude that the person lacks a corresponding moral belief. Our datum will then remain unexplained. On the other hand, if we do think that such cases are rare, we will naturally go on to infer that the person simply lacks a corresponding moral belief, and our datum will be explained.⁷¹

We saw in the last section that two popular versions of internalism that are weaker than the strongest version cannot explain one aspect of the practical character of morality, namely, the fact that we are generally reluctant to attribute a moral belief to someone who shows no sign of having a corresponding motivation. In this section, we have been asking whether there is a version of internalism weaker than the strongest version that *can* explain that datum. We have seen that more than one version of internalism fits this bill. Specifically, we saw that an indefeasible version of very strong internalism as well as a statistical form of very strong defeasible internalism can both explain our datum. Furthermore, we saw that other forms of very strong defeasible internalism can also explain the datum provided that we think that something (practical irrationality, the states included on the list of potential defeaters, or abnormality according to some salient standard of normality) is rare. I will now conclude this section by briefly returning to the idea that there is a conceptual connection between moral judgment and *overriding* motivation.

⁷¹ Dreier might think that there is a conceptual guarantee that it is rare for a case to be abnormal according to some salient standard of abnormality. This is suggested by his remark that the notion of normality he defends “is not *primarily* statistical.” Ibid., 7, my emphasis. If there is a statistical component to Dreier’s view, then it is subject to any general concerns one might have about the idea of a conceptual statistical truth.

Another version of internalism that is weaker than the strongest version is what I call “overriding defeasible internalism.” Like very strong internalism, overriding defeasible internalism has to my knowledge not been discussed in the literature. According to this version of internalism, it is a conceptual truth that *in normal cases* someone who has a moral belief has an overriding motivation to act as the belief prescribes. Overriding defeasible internalism thus asserts that there is conceptual connection between moral judgment and overriding motivation. In that respect, it resembles the strongest version of internalism. However, overriding defeasible internalism is weaker than the strongest version in that it grants the conceptual possibility of cases in which someone has a moral belief but lacks a corresponding overriding motivation. Indeed, it is consistent with the conceptual possibility of cases in which someone has no motivation whatsoever to act in accordance with one of her moral beliefs. So we should ask whether overriding defeasible internalism can explain our general reluctance to attribute a moral belief to someone who shows no sign of having a corresponding motivation. And at this point the answer should be obvious: it depends on what one says about the defeasibility conditions. If one takes a statistical approach, then overriding defeasible internalism can explain our datum. If, on the other hand, one takes a normative or brute list approach, or if one leaves the defeasibility conditions unarticulated as Dreier does, then this version of internalism can explain the datum if, but only if, we think that something (practical irrationality, the states included on the list of potential defeaters, or abnormality according to some salient standard of normality) is rare.

Conclusion

So internalists who reject the strongest version of internalism but who hope to derive support from the fact that we are generally reluctant to attribute a moral belief to someone who shows no sign of having a corresponding motivation have options. They can embrace very strong indefeasible internalism or a defeasible form of very strong or overriding internalism. However, if they choose one of the second two options, then they must either cash out the defeasibility conditions in statistical terms or be prepared to argue that some state (practical irrationality, whatever they include on the list of potential defeaters, or abnormality according to some salient standard of normality) is rare. Furthermore, these internalists cannot embrace two popular versions of internalism.

I have not attempted to assess the overall merits of any of the versions of internalism that can explain our datum, but it goes without saying that any of them would be controversial. Something that is perhaps worth mentioning is that nothing forces the internalist to try to derive support from their (putative) ability to explain our general reluctance to attribute a moral belief to someone who shows no sign of having a corresponding motivation, or for that matter any other aspect of the practical character of morality. Yet, it *is* often thought that internalism derives support from its ability to explain the practical character of morality, and I hope to have shown that while that thought may be true as applied to some versions of internalism and some aspects of the practical character of morality, some versions of internalism cannot explain at least one of the central ways in which morality is practical.

Chapter Four

The Frege-Geach Problem

Introduction

The topic of this chapter is the so-called Frege-Geach problem for non-cognitivism. There is no universally accepted account of what the Frege-Geach problem consists in, so I simply discuss four arguments against non-cognitivism that have all appeared in the literature under that heading. I have two goals. The first is to show that the arguments I present are distinct arguments. If this is right, then an adequate response to any one of these arguments may not be an adequate response to “the Frege-Geach problem,” since one or more of the other arguments may remain outstanding. This point is not always appreciated.⁷² The second goal is to show that, as they stand, none of these arguments provides sufficient reason to reject non-cognitivism. This point is admittedly not very controversial. Even philosophers who oppose non-cognitivism typically concede that the arguments against non-cognitivism that I discuss here require supplementation, and the non-cognitivist strategies that I sketch for responding to the arguments are original only in their presentation.⁷³ Yet, while my thesis is widely accepted, it is not usually thoroughly argued. My hope is that doing so in a context that carefully distinguishes the different arguments that have been discussed under the heading “the

⁷² Daniel Stoljar, for example, seems to overlook this point. Daniel Stoljar, “Emotivism and Truth Conditions,” *Philosophical Studies* 70 (April 1993). See also James Dreier, “Expressivist Embeddings and Minimalist Truth,” *Philosophical Studies* 83 (1996): 30-34.

⁷³ Bob Hale is an example of a critic of non-cognitivism who concedes that the usual presentation of the Frege-Geach problem is not decisive as it stands. See Bob Hale, “Can There Be a Logic of Attitudes?” in *Reality, Representation, and Projection*, ed. John Haldane and Crispin Wright (New York: Oxford University Press, 1993), 337-341.

Frege-Geach problem” will make perspicuous exactly what challenges “the Frege-Geach problem” raises for non-cognitivism. I begin by explaining how I am understanding non-cognitivism.

Non-cognitivism

Non-cognitivism is sometimes presented as a view about moral thought, a view about what it is to think that, say, lying is morally wrong. Thus understood, non-cognitivism is the view that the states of mind that we ordinarily call “moral beliefs” are not *really* beliefs at all but are rather attitudes of approval and disapproval. On this view, to think that lying is wrong, for example, is not to think that things are a certain way, that lying has, in addition to whatever other properties it has, the property of being wrong. Rather, it is just to disapprove of lying. Non-cognitivism is also often presented as a view about moral discourse. Consider the sentence “the door is closed.” There is an important sense in which this sentence has a fact-stating role. We use it to describe how things are: That is what the sentence is *for*. Now consider the expression “Ouch!” There is an important sense in which this expression does *not* play a fact-stating role. We do not ordinarily use “Ouch!” to describe how things are. Rather, we use it give voice to pain. As a view about moral discourse, non-cognitivism holds that, despite appearances, moral sentences are like “Ouch!” and unlike “the door is closed” in that they do play a fact-stating role. According to the non-cognitivist, just as we use “Ouch!” to give voice to pain, we use moral sentences to voice attitudes of approval and disapproval. In a standard terminology, non-cognitivists claim that the *semantic function* of moral sentences is to *express* attitudes, where *expressing an attitude* is distinct from *saying that one has that*

attitude.⁷⁴ (Compare how uttering “Ouch!” is different from uttering “I am in pain.”) Thus, on this view, someone who sincerely utters the sentence “lying is wrong” is not trying to state a fact about lying, not trying to assert that lying has, in addition to whatever other properties it has, the property of being wrong. Rather, this person is merely expressing disapproval of lying.⁷⁵

The arguments that I will discuss can, I think, be applied in different forms to both non-cognitivism about moral thought and non-cognitivism about moral discourse.⁷⁶ They are usually presented, however, as problems for non-cognitivism about moral discourse, and that is how I will treat them here. By “non-cognitivism,” then, I mean the view that the semantic function of moral sentences is not to state facts but is to express attitudes of approval and disapproval.

I want to make one more clarificatory point about how I am understanding non-cognitivism before I turn to the arguments discussed under the heading “the Frege-Geach problem.” I am following a standard practice in defining non-cognitivism as the view that the semantic function of moral sentences is not to state facts but to express attitudes of approval and disapproval. But it is important to keep in mind that “semantic” is being used here in a broad sense. Specifically, the notion of “semantic function” at work refers to the role that a form of words plays in our overall lives. In this sense of “semantic function,” to specify the semantic function of an expression is to explain what the expression is *for*; it is to explain, as Simon Blackburn puts it, *what we are up to* when we use the expression.⁷⁷ Now, this sense of “semantic function” is admittedly not as clear as

⁷⁴ See Miller, “Introduction to Metaethics,” 2.

⁷⁵ See p. 1-2.

⁷⁶ See p. 5-24.

⁷⁷ See Blackburn, *Spreading the Word*, 192.

one would like. Does it refer to the *evolutionary* purpose of the expression or to a *speaker's* purpose for using it? If the latter, are we concerned in the first instance with individual speakers or with language communities? Unfortunately, philosophers writing about non-cognitivism have not given these sorts of questions the attention that they deserve. For our purposes, however, we can leave this broad notion of “semantic function” relatively rough. It will be enough to keep in mind the kind of examples introduced above. Indeed, we can even think of the phrase “the semantic function of moral sentences is not to state facts but to express attitudes of approval and disapproval” as being simply shorthand for the following: “in the same sense (whatever it is) in which we do not ordinarily use “Ouch!” to state a fact but rather to give voice to pain, we do not ordinarily use moral sentences to state facts but rather to give voice to attitudes of approval and disapproval.”

I turn now to four arguments that have been discussed under the heading “the Frege-Geach problem.” To reiterate, my goals are to show how the arguments differ from one another and to show that, as they stand, they do not provide sufficient reason to reject non-cognitivism.

The Simple Argument

Discussions of the Frege-Geach problem often begin with what I call “the simple argument.” The simple argument turns on the idea that moral expressions can appear meaningfully in embedded contexts. For example, they can appear meaningfully as antecedents and consequents of conditionals. Consider again “lying is wrong.” According to the non-cognitivist, the semantic function of “lying is wrong” is to express disapproval of lying. But now consider the following sentence:

1. If lying is wrong, getting your little brother to lie is wrong.

The expression ‘lying is wrong’ clearly has a semantic function in (1). It also seems clear, however, that its function in (1) is *not* to express disapproval of lying. So even if we grant that the semantic function of moral expressions in *unembedded* contexts is to express attitudes of approval and disapproval, non-cognitivism seems to give the wrong analysis of ‘lying is wrong’ as it appears in this embedded context.⁷⁸

As James Dreier points out, this argument can be read in both a strong and a weak way.⁷⁹ Read in the strong way, the argument tries to refute non-cognitivism by counterexample: If non-cognitivism is true, then the semantic function of “lying is wrong” in (1) is to express disapproval of lying; but that is not its function in (1); therefore, non-cognitivism is false. The problem with this strong argument is that non-cognitivists should be allowed the opportunity to revise their thesis in light of the charge that they give wrong analyses of embedded moral sentences. This is in fact what non-cognitivists have done. Non-cognitivists typically respond to the simple argument by restricting the claim that the function of moral sentences is to express attitudes of approval and disapproval to unembedded contexts. They then attempt to supplement this restricted claim with claims about how moral sentences function in various embedded contexts.⁸⁰ Now, it *may* be that this strategy cannot succeed. But that is something that needs to be shown. The immediate point is that the strong version of the argument cuts off the debate too soon.

⁷⁸ Peter Geach, “Assertion,” *The Philosophical Review* 74 (October 1965): 461-465; Miller, “Introduction to Metaethics,” 40-42.

⁷⁹ Dreier, “Expressivist Embeddings,” 31.

⁸⁰ See, for example, Blackburn, *Spreading the Word*, 189-196; Gibbard, “Wise Choices,” 83-102.

The weak version of the simple argument tries to show not that non-cognitivism is *false* but rather that it is *incomplete*: Non-cognitivism has at most explained the semantic function of moral sentences in unembedded contexts; but as sentences such as (1) make clear, moral sentences also have a semantic function in embedded contexts; therefore, non-cognitivism is at best incomplete. Understood in this way the simple argument does not cut off the debate too soon. In fact, it issues a challenge to the non-cognitivist, a challenge that non-cognitivists have actually taken up. Again, it may be that the non-cognitivist will not ultimately be able to give adequate analyses of embedded moral sentences: the next argument we will look at tries to show just that. But this much at least is already clear: whether it is read in the strong or weak way, the simple argument does not by itself give us reason to reject non-cognitivism.

The Equivocation Argument

From here forward I will incorporate the revision forced by the simple argument and take non-cognitivism to be the view that *unembedded* moral sentences function to express attitudes of approval and disapproval. The next argument that I want to discuss, which I call “the equivocation argument,” aims to show that even this more restricted version of non-cognitivism is false.⁸¹ It can be stated as follows. Non-cognitivists claim that the function of “lying is wrong” in unembedded contexts is to express disapproval of lying. But now consider the following argument:

⁸¹ My presentation of this argument closely follows Miller’s in Miller, “Introduction to Metaethics, 40-41. The inference that equivocates on “head” is his.

1. If lying is wrong, getting your little brother to lie is wrong.
2. Lying is wrong.

3. Therefore, getting your little brother to lie is wrong.

Call this argument “L.” Since “lying is wrong” appears unembedded in (2), non-cognitivism implies that it is functioning there to express disapproval of lying. But that is clearly *not* the function of “lying is wrong” in (1). So for their position to be plausible, non-cognitivists have to claim that “lying is wrong” *has different semantic functions in (1) and (2)*. But that means that non-cognitivists have to deny that L is valid, since if “lying is wrong” has different semantic functions in (1) and (2), then L is guilty of the fallacy of equivocation and is thus on a par with the following obviously invalid argument:

4. If something has a head on it, it has eyes and ears.
5. My beer has a head on it.

6. Therefore, my beer has eyes and ears.

Call this argument “H.” Unlike L, H *is* valid. So, “lying is wrong” does *not* have different semantic functions in (1) and (2). So, non-cognitivism is false.

Like the simple argument, the equivocation argument depends crucially on the point that moral sentences appear meaningfully in embedded contexts. It is, however, a distinct argument. The non-cognitivist can sidestep the simple argument by simply restricting her central claim to embedded moral sentences. But the equivocation argument cannot be dealt with so easily.

I will now show that this argument is not conclusive. First some set up. The equivocation argument depends crucially on the idea that if “lying is wrong” does not have the same semantic function in (1) and (2), then L is no better than paradigmatic cases of the fallacy of equivocation like H. But now consider the following argument:

7. If it is raining, then the pavement is wet.

8. It is raining.

9. Therefore, the pavement is wet.

Call this argument “R.” R, like H, has the very same surface form as L. Also, just as “lying is wrong” has different semantic functions in (1) and (2), the occurrences of “it is raining” in (7) and (8) seem to have different semantic functions in the broad sense of “semantic function” we have been using. Specifically, “it is raining” plays a fact-stating role in (8) but not in (7).⁸² Yet nobody thinks that this implies that R is invalid in virtue of committing the fallacy of equivocation. Why not? Alexander Miller gives the standard explanation:

Why is this non-moral case of *modus ponens* not invalid in virtue of the fact that ‘It is raining’ is asserted in ... [(8)], but not in ... [(7)]? The answer is of course that the state of affairs asserted to obtain by ‘It is raining’ in ... [(8)] is the same as that merely hypothesized to obtain in ... [(7)]. In ... [(8)] ‘It is raining’ is used to assert that a state of affairs obtains (it’s raining), and in ... [(7)] it is asserted that, if that state of affairs obtains, so does another (the streets being wet). Throughout, the semantic function of the sentences concerned is given in terms of the states of affairs asserted to obtain in simple assertoric contexts.⁸³

The idea is that R does not commit the fallacy of equivocation because ‘it is raining’ *does* in fact have the same semantic function in (7) and (8), namely to represent a certain state

⁸² Ibid., 42.

⁸³ Ibid.

of affairs. Of course, that state of affairs is represented as actually obtaining in (8) and represented only as a mere “hypothesis” in (7), but that is not a difference in *semantic function*. Miller’s explanation also indicates why H *does* commit the fallacy of equivocation: “has a head on it” *does not* represent the same state of affairs in H’s first and second premises.

Crucially, Miller’s explanation employs a *narrower* conception of semantic function than the one we have been using. According to this narrower conception, the semantic function of a sentence consists in its representing a certain state of affairs, namely the state of affairs that it asserts to obtain in “simple assertoric contexts.” We, recall, have been using “semantic function” to refer to a much broader property of certain expressions, namely to the role that the expression plays in our overall lives, or, as Blackburn puts it, to “what we are up to” when we use the expression. “What we are up to” when we utter “it is raining” in “simple assertoric contexts” is clearly different from “what we are up to” when we utter that expression in the context of an antecedent of a conditional. I have no interest here in entering into a debate on which, if either, of these conceptions of semantic function is the *right* one. Rather, I want to show how the non-cognitivist can use this distinction to show that the equivocation argument is inconclusive.⁸⁴

According to the equivocation argument, L commits the fallacy of equivocation, and is hence invalid, if “lying is wrong” has different semantic functions in (1) and (2). Given the two conceptions of semantic function we have just distinguished, a natural

⁸⁴ The argument that follows has been influenced by remarks by Bob Hale. Hale, “Logic of Attitudes, 338-339.

question to ask is whether the proponent of the equivocation argument is employing the narrow or broad conception of semantic function. Let's look at both answers.

Suppose first that the proponent of the equivocation argument has in mind the narrow conception of semantic function. His claim would then be that L commits the fallacy of equivocation, and is hence invalid, if "lying is wrong" does not represent the same state of affairs in (1) and (2). The non-cognitivist has two strategies available at this point.

First, she can revise her thesis once again. Specifically, instead of claiming that *the* semantic function of unembedded moral sentences is to express attitudes of approval and disapproval she can claim that that is their *primary* semantic function. The non-cognitivist can then say that a *secondary* semantic function of "lying is wrong" in both (1) and (2) is to represent a certain state of affairs.⁸⁵

Alternatively, the non-cognitivist can concede that "lying is wrong" has different semantic functions in (1) and (2) in the narrow sense of "semantic function" but deny that this implies that L is invalid. This strategy rests on showing that L can be valid in virtue of (inter alia) facts about the semantic function of "lying is wrong" in (1) and (2) *in the broad sense of "semantic function."* Simon Blackburn, for example, takes this tack. According to Blackburn, L is valid because not accepting L's conclusion "clashes" with "what one is up to" when one accepts (1) and (2). To accept (1), according to Blackburn, is to express disapproval of moral sensibilities that disapprove of lying but do not disapprove of getting your little brother to lie. On this analysis, someone who accepts (1) and (2) but not (3) has a combination of attitudes that he himself disapproves of: He has a "clash of attitudes" issuing from a "fractured sensibility." Of course, it *may* be that the

⁸⁵ van Roojen, "Moral Cognitivism."

validity of L cannot be explained in this way and that L's validity depends crucially on "lying is wrong" representing the same state of affairs in (1) and (2). But that cannot be taken for granted.⁸⁶

At this point the proponent of the equivocation argument might remind the non-cognitivist of the contrast between R and H and say something like the following. H has the same surface form as R, but R is valid while H is obviously invalid. *Something* must account for this difference, and the only plausible hypothesis is that the difference consists in the fact that in H, but not in R, the same expression does not represent the same state of affairs in the antecedent of the conditional premise and in the non-conditional premise. L is just like H in this respect. So, the objector will ask, what principled reason is there for taking seriously the possibility that L can avoid H's fate and be valid even if "lying is wrong" does not represent the same state of affairs in L's two premises?

The non-cognitivist can, I think, answer this challenge. He need only deny the claim that the *only* plausible explanation for H being invalid while R is valid is that "has a head" represents the same state of affairs in H's premises while "it is raining" does not represent the same state of affairs in R's premises. Indeed, the non-cognitivist response we are now considering suggests another possible explanation: there is a connection between "what we are up to" when we accept (7) and "what we are up to" when we accept (8) that renders L valid, but there is no connection between "what we are up to" when we accept (4) and "what we are up to" when we accept (5) that renders H valid.

I conclude, then, that the equivocation argument is not conclusive if it is employing the narrow conception of semantic function. Suppose, then, that the proponent

⁸⁶ Hale, "Logic of Attitudes, 338-339; Blackburn, *Spreading the Word*, 189-196.

of the equivocation argument has in mind the broader conception of semantic function. Read this way, the equivocation argument cannot pose a special problem for non-cognitivism, or so I will now argue.

It is clear that “lying is wrong” does not have the same semantic function in L’s two premises *in the broad sense of “semantic function.”* It plays a fact stating role in (1) but not in (2). Now consider what this shows: When “semantic function” is being used in the broad sense, “lying is wrong” will have different semantic functions in L’s two premises on *any* plausible analysis of “lying is wrong.” That means that if the proponent of the equivocation argument is correct, L is invalid in virtue of committing the fallacy of equivocation—as for that matter is R—on *any* plausible analysis of “lying is wrong.” It follows that when “semantic function” is understood in the broad sense of constituting the role an expression plays in our overall lives, “what we are up to” when we use the expression, the claim that L is invalid in virtue of committing the fallacy of equivocation cannot pose a special problem for non-cognitivists.

The Standard Validity Argument

The next argument I want to look at, which I call “the standard validity argument,” also turns on the idea that non-cognitivism is incompatible with the validity of certain arguments involving moral sentences. It runs as follows. Consider L. According to non-cognitivism, (2) and (3) function to express attitudes of disapproval. But expressions of disapproval are not truth-apt; that is, they can be neither true nor false. (Compare how “Ouch!” can be neither true nor false.) So non-cognitivism implies that L’s second premise and conclusion are not truth-apt. Now, for an argument to be valid is for it to be such that it is impossible for its premises to be true and its conclusion false. So

the valid/invalid distinction applies only to arguments whose premises and conclusions are truth-apt. It follows that non-cognitivism implies that L is not valid. But L *is* valid. So non-cognitivism is false.⁸⁷

Note that this argument is different from the equivocation argument. Though both arguments rely on the thought that arguments involving moral sentences can be valid, only the equivocation argument relies on the fact that moral expressions can appear meaningfully in embedded contexts. The thrust of the standard validity argument is that non-cognitivism is inconsistent with the validity of certain apparently valid arguments because it implies that elements of those arguments are not truth-apt. To be sure, I *presented* the standard argument in a way that appealed to the apparent validity of an argument in which moral sentences do occur in embedded contexts, but that was not essential. I could just as easily have appealed to an apparently valid argument that includes only unembedded moral sentences, such as

2. Lying is wrong.

2. Therefore, lying is wrong.

or

2. Lying is wrong.

10. Therefore, it is raining or it is not the case that it is raining.

As with the equivocation argument, there are at least two strategies that the non-cognitivist can deploy in response to this argument. First, she again has the option of

⁸⁷ See Frank Jackson, "Non-cognitivism, Validity, and Conditionals," in *Singer and His Critics*, ed. Dale Jamieson (Oxford: Blackwell Publishers, 1999): 18-19; Dreier, "Expressivist Embeddings," 34.

revising her central claim so that it says only that the *primary* function of unembedded moral sentences is to express attitudes of approval and disapproval. She can then argue that unembedded moral sentences are truth-apt in virtue of some *secondary* semantic function.⁸⁸

Alternatively, as Bob Hale and others have pointed out, the non-cognitivist's opponent cannot just assume from the outset that necessary truth preservation is the only legitimate conception of validity.⁸⁹ The non-cognitivist has the option of arguing for an alternative, non-truth-theoretic conception. The non-cognitivist can, for example, endorse Blackburn's "clash of attitudes" conception of validity.

A proponent of the standard validity argument may object that it is ad hoc to appeal to the possibility that L is valid in some non-standard sense of validity. There are, however, independent reasons for being open to non-standard conceptions of validity. For example, it may be that some non-standard account of validity simply does a better job of systematizing our pretheoretic judgments concerning which arguments are, and which arguments are not, valid. Some philosophers think that this is the case, for example, with the view that, at least when it comes to arguments involving conditionals, validity consists not in the necessary preservation of *truth*, but rather in the necessary preservation of *intuitive probabilities*.⁹⁰

The proponent of the standard validity argument might also object that any appeal to a non-standard view of validity can succeed only in changing the subject.⁹¹ More precisely, he may argue as follows. Our intuitive reaction to L is that it is valid *in the very*

⁸⁸ van Roojen. "Moral Cognitivism."

⁸⁹ Hale, "Logic of Attitudes," 339. Max Kölbel, *Truth Without Objectivity* (London: Routledge, 2002), 134.

⁹⁰ Jackson, "Non-cognitivism," 22.

⁹¹ Jackson, "Non-cognitivism," 20.

same sense that R is valid. We think that R is valid in the standard sense of being necessarily truth preserving. Thus, our intuitive reaction to L is that it too is valid in the sense of being necessarily truth preserving. *That* is the intuition that the non-cognitivist needs to accommodate. He cannot do so by showing that L is valid in *some other sense.* That would just change the subject.

This worry is understandable, but it too can be answered. The first thing for the non-cognitivist to point out is that we can say illuminating things about validity without (at least overtly) invoking truth. For example, we know that valid arguments are, in some sense of “good,” good inferences. Furthermore, we know that valid arguments are not the only arguments that are good in that sense: certain other arguments, namely inductively strong arguments, are good in the same sense of “good” in which valid arguments are good. Together these two considerations give us a bead on validity that does not (at least overtly) presuppose that only arguments composed exclusively of truth-apt elements can be valid.

Furthermore, the non-cognitivist can point out that we can guard against the subject being changed by adopting certain adequacy conditions on any proposed non-standard account of validity. For example, Frank Jackson proposes that the subject is not changed if the non-standard account meets the following three constraints: the meaning constraint, the significance constraint, and the intuition constraint. The meaning constraint demands that the non-standard account “advert to a property intimately connected with the meanings of the sentences involved.” The significance constraint demands that the account be such that it is transparent why validity is a desirable

property. The intuition constraint demands that all intuitively valid arguments come out as valid on the account.⁹²

At this point the objector may complain that the non-cognitivist has simply ignored the argument he advanced in favor of his claim that appealing to a non-standard view of validity is bound to result in a changing of the subject. That argument turned on the idea that our intuitive reaction to L is that it is valid *in the very same sense* that R is valid. Since we think that R is valid in the standard sense of being necessarily truth preserving, the argument continues, it follows that our intuitive reaction to L is that it too is valid in that standard sense. The objector will insist that the non-cognitivist offer a direct response to this argument.

I think he can. The right response is that it is just not obvious that when we judge R to be valid we are *thereby* judging the argument to be necessarily truth preserving. If that's right, then it is also not obvious that in having the intuition that L is valid in the same sense that R is valid we are *thereby* having the intuition that L is valid in the sense of being necessarily truth preserving. Of course, we *do* judge R to be valid, *and* we *do* judge it to be necessarily truth preserving, but the objector needs more than this. He needs the premise that *at a level of judgment individuation fit to sustain his worries about the subject being changed* judging that R is necessarily truth preserving *is part of what one does* when one judges that R is valid. That is what is not obvious. For all the above argument shows, the intuition that L is valid in the very same sense that R is valid is just the intuition that the former, like the latter, is valid in the very neutral sense sketched above of being a good, non-inductive inference.

⁹² Jackson, "Non-cognitivism," 20-21.

The Shared Principle Argument

I call the final argument that I want to discuss the “shared principle argument.” This argument starts from the premise that L follows the same inference principle as straightforward cases of modus ponens like R. Since R is a case of modus ponens, the argument continues, it follows that L is also a modus ponens argument. But L is a case of modus ponens only if “lying is wrong” has the same semantic function in its first and second premises, and we have seen that that is not the case if non-cognitivism is true. Therefore, non-cognitivism is false.⁹³

Note that the shared principle argument differs from both the equivocation argument and the standard validity argument in at least two important respects. First, unlike those previous two arguments, it is not a premise of the shared principle argument that non-cognitivism is incompatible with L’s being valid.⁹⁴ The shared principle argument claims only that non-cognitivism is incompatible with a certain *explanation* for L’s validity, namely its being a case of modus ponens. Second, of the three arguments, the shared principle argument is also the only one that relies on the premise that L and inferences like R have the same logical form. For all the previous two arguments claim, R and L are both valid but only R is a case of modus ponens.

This argument is also inconclusive.⁹⁵ The non-cognitivist again has at least two things to say. First, she can point out that if the first premise of the shared principle argument is referring to the “deep forms” of L and R, we need an argument for that premise. For while it is obvious that L and R share the same surface form—that is,

⁹³ Susan Hurley, “Frege, the Proliferation of Force, and Non-cognitivism,” *Mind* 93 (October 1984).

⁹⁴ *Ibid.*, 574.

⁹⁵ The following argument is based on remarks that Blackburn makes in Simon Blackburn, “Attitudes and Contents,” *Ethics* 98 (April 1988): 501-502.

obvious that they *look* like the same sort of argument—it is not obvious that their underlying logics are the same.

Second, notice that the proponent of the shared principle argument *must* be using “modus ponens” in some “deep” sense, since if to be a case of modus ponens is just to *look* like R, there is no reason to think that non-cognitivism is inconsistent with L’s being a case of modus ponens. Now, there is no universally accepted account of what it takes for an argument to have the “deep form” of modus ponens. The non-cognitivist can thus point out that it is a theoretically live option that, contrary to appearances, L is not a case of modus ponens. Furthermore, if it does turn out that L is a case of modus ponens in some suitably “deep” sense, we cannot say whether that fact is in conflict with non-cognitivism until we settle the issue of what that “deep form” consists in.

Conclusion

I have here distinguished four arguments against non-cognitivism that have all appeared in the literature in connection with the so-called Frege-Geach problem and have shown that they are all inconclusive. These arguments do nevertheless pose challenges for non-cognitivism. Specifically, the non-cognitivist needs to (a) provide plausible analyses of moral sentences in various embedded contexts, (b) show that the fact that non-cognitivism apparently gives moral sentences different semantic functions in arguments like L is consistent with the apparently validity of such arguments, (c) show that the fact that non-cognitivism apparently renders unembedded moral sentences incapable of being true or false is consistent with such sentences figuring in apparently valid arguments, and (d) show that the fact that non-cognitivism apparently gives moral sentences different semantic functions in arguments like L is consistent with such

arguments apparently following the same inference principle as arguments like L. I have only tried to briefly sketch ways in which the non-cognitivist can attempt to meet these challenges.

Chapter Five

Why Emotivists Shouldn't Love Inconsistency

Introduction

In “Why Emotivists Love Inconsistency,” Gunnar Björnsson writes:

An important fact about our moral thinking is that we give up moral opinions to avoid what we naturally think of as inconsistencies, and we come to have new moral opinions as a result of having other opinions from which the new ones seem to follow. Critics have seen this as undermining emotivism in a decisive way: such a practice would make sense only if moral opinions are beliefs—states capable of truth and falsity and having the function to guide us to our goals.⁹⁶

Björnsson dubs this the “problem of reasoning” for emotivists and proposes a solution.⁹⁷

In this chapter I identify two problems with that solution. I begin by clarifying how Björnsson understands “emotivism” and what he takes the problem of reasoning to be. I then present Björnsson’s solution to the problem. Finally, I state my two objections to the solution.

The Problem of Reasoning

Unfortunately, Björnsson is not very clear about what the problem of reasoning amounts to. This much, however, is certain: it concerns the emotivist’s ability to explain the “important fact” that Björnsson describes in the above passage. So we need to begin by asking three questions. First, what is emotivism? Second, what exactly is this “important fact”? And third, why might one think that emotivists cannot adequately explain the “important fact”? I will address these questions in turn.

⁹⁶ G. Björnsson, “Why Emotivists Love Inconsistency,” *Philosophical Studies* 104 (2001): 85.

⁹⁷ *Ibid.*

Björnsson characterizes emotivism as the view that “moral opinions are wishes and desires and that the function of moral language is to ‘express’ such states’.”⁹⁸ Understood this way, emotivism consists of both a psychological thesis and a linguistic thesis, a thesis about the nature of certain mental states and a thesis about the function of certain bits of language. It soon becomes clear, however, that Björnsson is primarily concerned with the psychological thesis. In fact, questions about the function of moral language do not seem to play a role in either Björnsson’s understanding of the problem of reasoning or his solution to it. For our purposes, then, we can ignore the linguistic thesis and think of emotivism as just the psychological thesis that to have a moral opinion is not to have a belief about how things are but is rather to have a certain wish or desire.⁹⁹

⁹⁸ Ibid., 81.

⁹⁹ There is a complication here. Björnsson begins his article: “Emotivists hold that moral opinions are wishes and desires, and that the function of moral language is to ‘express’ such states.” Ibid. However, he almost immediately replaces “desire” with “optation”: “What makes a theory about the workings of moral opinions emotivist are certain theoretical *identifications* of certain kinds of moral opinions and kinds of wishes or desires, or, as I shall put it, moral ‘optations’.” Ibid., 82, original emphasis. It is unclear why he does this. More importantly, it is unclear what he means by “optation.” At times he writes as if optations are a species of desire or wish. He says, for example: “Exactly how to distinguish moral optations from other kinds of wishes and desires is a difficult matter.” Ibid. At other times, however, he apparently wants to leave open the possibility that moral optations are a kind of belief-desire hybrid or even simply *beliefs!* Ibid., 87. Consider: “This means that even though I refrain from defining emotivism negatively as the view that moral opinions are not beliefs, our problem remains.” Ibid., 87. Relatedly, Björnsson seems to waffle on the issue of whether emotivism implies that moral opinions are not truth-apt. Compare the last sentence of the quotation that begins the body of the present chapter with something Björnsson says in the conclusion: “... the literature contains a number of sophisticated efforts to show that plausible emotivist accounts imply that moral opinions can be true or false. But this is not the place to decide whether the above account suggests cognitivism, or whether such a cognitivism would be of much interest. Our focus has been elsewhere. What we have seen is how emotivism (and indeed cognitivism) can begin to explain our intuitions of inconsistency.” Ibid., 103. The aforementioned quotations are odd, to say the least, if we are leaving open the possibility that emotivism is consistent with moral opinions’ being truth-apt. And if we are also leaving open the possibility that moral optations are beliefs, it is hard to see what the contrast between emotivism and cognitivism is supposed to be. For what it’s worth, I suspect that Björnsson is jumping back and forth among three related but distinct issues: (i) Are moral opinions necessarily motivating? (ii) Are moral opinions beliefs or desires? (iii) Are moral opinions truth-apt? In any event, I will follow Björnsson’s original characterization of emotivism as the view that moral opinions are kinds of wishes or desires. I will further assume—as Björnsson apparently does throughout most of his article—that it is part of this view that moral opinions are not truth-apt. I feel justified in doing this because the alternative leaves Björnsson’s paper highly obscure. For clarity and convenience, I replace “optation” with “desire” in a couple of quotations. I signal each such replacement.

I turn now to the second question. The problem of reasoning is that the emotivist seems unable to make sense of an “important fact about our moral thinking.” But what exactly is this “important fact”? Recall how Björnsson puts it: “we give up moral opinions to avoid what we naturally think of as inconsistencies, and we come to have new moral opinions as a result of having other opinions from which the new ones seem to follow.” This remark can be broken down into two related points.¹⁰⁰ The first is that moral opinions enter into relations of inconsistency and consequence, or at least relations that we naturally think of as inconsistency and consequence. Put another way, moral opinions at a minimum *seem* to stand in logical relations. Björnsson illustrates this point with an example. Consider the following two sets of opinions:

1.
 - A. If lying makes one nervous then telling one’s little brother to lie makes one nervous.
 - B. Lying makes one nervous.
 - C. Telling one’s little brother to lie does not make one nervous.

2.
 - A’. If lying is wrong then telling one’s little brother to lie is wrong.
 - B’. Lying is wrong.
 - C’. Telling one’s little brother to lie is not wrong.¹⁰¹

As Björnsson points out, (2) seems no less inconsistent than (1). Similarly, just as

- D. Telling one’s little brother to lie makes one nervous.¹⁰²

seems to follow from (A) and (B),

¹⁰⁰ Brian Kierland helped me to see this. In an earlier draft I had simply identified the “important fact,” with the second of the two points.

¹⁰¹ These come straight from Björnsson with one modification. *Ibid.*, 85-86. Björnsson uses the letters to stand for sentences, not opinions. He later introduces the device of adding asterisks to pick out the opinions expressed by the sentences that the letters without the asterisks pick out. Thus, he uses “A*” to stand for the opinion that if lying makes one nervous, getting one’s little brother to lie makes one nervous; “B*” to stand for the opinion that lying makes one nervous; and so on. Nothing I say here depends on keeping sentences neatly separated from the opinions they express, so I am using the letters without the asterisks to stand for the opinions directly. I modify a quote below to accommodate this. I signal that modification.

¹⁰² *Ibid.*, 88.

D'. Telling one's little brother to lie is wrong.

seems to follow from (A') and (B'). The second point is that we will give up an opinion to avoid accepting (what we naturally think of as) an inconsistency that involves a moral opinion, and we may acquire a new opinion as a result of (what we naturally think of as) an implication involving a moral opinion. For example, we are just as averse to accepting every opinion in (2) as we are to accepting every opinion in (1). And we are just as happy to infer (D') from (A') and (B') as we are to infer (D) from (A) and (B).

There are, then, two points that make up the "important fact": moral opinions stand in (what we naturally think of as) relations of inconsistency and consequence, and we sometimes adjust our opinions in accordance with (what we naturally think of as) inconsistencies and implications involving moral opinions. The question now is why one might think emotivists cannot adequately explain one or both of these points. Again, Björnsson is not as clear as one would like, but he seems to have in mind at least three worries.

The first worry corresponds to the first point contained in the "important fact."¹⁰³ The most natural explanation of why moral opinions *seem* to enter into relations of inconsistency and consequence is that they *really do* enter into these relations. But this explanation is available to the emotivist only if she is willing to abandon the standard conceptions of inconsistency and consequence. This is because the standard conceptions are truth-theoretic, and emotivism implies that moral opinions are not capable of being true or false. On the standard conceptions, a set of opinions is inconsistent just when it is impossible for all of its members to be true, and an opinion, *o*, follows from a set of opinions, *O*, just when it is impossible for *o* to be false when all of the members of *O* are

¹⁰³ Brian Kierland helped me to see this.

true.¹⁰⁴ Given these conceptions, moral opinions can stand in relations of inconsistency and consequence only if they are capable of being true or false. Yet, if emotivism is true, moral opinions are desires. And while it can be true or false *that someone has* a certain desire, no desire can *itself* be true or false. So given the standard conceptions of inconsistency and consequence, moral opinions are not the right sort of thing to stand in those relations if emotivism is true. The fact that moral opinions nevertheless seem to stand in relations of inconsistency and consequence forces the emotivist to do one of two things: she can either argue that we should reject the standard conceptions of inconsistency and consequence in favor of some non-truth-theoretic conceptions, or she can abandon the most natural explanation of why moral opinions seem to stand in relations of inconsistency and consequence and try to explain this in a way that does not advert to moral opinion's *really* entering into those relations. This seems problematic for the emotivist because we presumably would prefer to retain *both* the standard conceptions of inconsistency and consequence *and* the idea that moral opinions seem to enter into those relations because they really do enter into them. Insofar as cognitivism—the view that moral opinions are beliefs—allows us to say both of these things, it enjoys an apparent advantage over emotivism.

The second worry concerns both of the two points that comprise the “important fact.” It is a worry about generality. Corresponding to the first point, the fact that moral opinions seem to enter into relations of inconsistency and consequence, is the fact that non-moral opinions that are paradigmatic beliefs seem to enter into such relations. ((A), (B), (C), and (D) are, of course, examples of such opinions.) Now, if cognitivism is true and moral opinions are beliefs, then a single phenomenon underlies both of these facts:

¹⁰⁴ Ibid., 86.

beliefs seem to enter into relations of consistency and inconsistency. Cognitivists can thus explain the point about moral opinions and the corresponding fact about non-moral opinions in one fell swoop. Any explanation for the point about moral opinions will also explain the corresponding fact about non-moral opinions, and vice versa. For emotivists, however, explaining why *beliefs* seem to enter into relations of inconsistency and consequence will not also explain why moral opinions seem to enter into such relations. To explain *that* emotivists need to *also* explain why *desires* seem to enter into such relations. This looks worrisome for the emotivist because, presumably, having a single general explanation of why opinions seem to enter into relations of inconsistency and consequence is preferable to having one account covering non-moral opinions and another covering moral opinions.

Consider now the second point comprising the “important fact”: we adjust our opinions in light of (what we naturally think of as) inconsistencies and implications involving moral opinions. Corresponding to this point is the fact that we adjust our opinions in light of inconsistencies and implications involving only non-moral opinions that are paradigmatic beliefs. (For example, we will adjust our opinions to avoid accepting every opinion in (1), and we will infer (D) from (A) and (B).) As with the first point comprising the “important fact” and the corresponding fact about non-moral opinions, if cognitivism is true, a single phenomenon is at work. In this case it is that we revise our opinions in light of inconsistencies and implications among beliefs. Once the cognitivist explains this phenomenon, she will have explained both the second point comprising the “important fact” and the corresponding fact about inconsistencies and implications involving only non-moral opinions. But the emotivist is again forced to

provide two separate explanations: one for why we respond as we do to inconsistencies and implications involving only beliefs, and another for why we respond as we do to (what we naturally think of as) inconsistencies and implications that involve desires. And this looks troubling for the emotivist, since, presumably, we would again prefer to have the kind of general explanation the cognitivist can offer.¹⁰⁵

The third worry also concerns both points that comprise the “important fact.” It is that it is not clear whether we can even make sense of complex opinions like (A’) and (C’) if emotivism is true and moral opinions are desires.¹⁰⁶

We can now state the problem of reasoning. The problem is showing that someone who claims that moral opinions are desires as opposed to beliefs, and thus denies that moral opinions are capable of being true or false, can adequately explain the fact that moral opinions seem to enter into relations of inconsistency and consequence and the fact that we respond to these relations in the way that we do despite the three worries just discussed. I turn now to Björnsson’s attempt to show that this can, in fact, be done.

Björnsson’s Solution

Björnsson deploys a two-part strategy which he glosses thus:

Basically, the idea is this. We start with everyday experiences of the role that logical intuitions and complex opinions play in our thinking, and hypothesize that these roles are purposive, chosen by natural selection and our learning mechanisms We then ask whether and why we should expect to have such states given that our simple moral opinions are beliefs and desires, respectively If we come up with plausible answers with respect to beliefs, our initial hypotheses are vindicated. If we come up with plausible answers with respect to desires too, emotivism is vindicated. [More precisely: if we come up with plausible answers with respect to

¹⁰⁵ See *ibid.*, 87-88, 90.

¹⁰⁶ See *ibid.*, 81.

desires too, then we will have thereby shown that the emotivist can adequately explain the “important fact” despite the three worries.].¹⁰⁷

It is important to see that this is a *functionalist* strategy in two senses. First, it is functionalist in that it appeals to the roles that logical intuitions (specifically, intuitions of inconsistency and consequence) and logically complex opinions (for example, conditional opinions) play in our overall thinking. Second, the strategy is also functionalist in a teleological sense, since it involves taking the identified roles to be “purposive, chosen by natural selection and our learning mechanisms.” To anticipate, the first problem I identify for Björnsson’s solution concerns his account of the role that intuitions of inconsistency play in our overall thinking, while the second problem concerns his account of the teleological function of these intuitions. Both problems arise even when considering only paradigmatic beliefs. Part of my response to the above quote, then, is that, unless Björnsson can solve these problems, he has failed to even “come up with plausible answers with respect to beliefs.”

Björnsson begins by giving his analyses of logically complex opinions. They are as follows:

Negative opinions—opinions to the effect that something isn’t so-and-so—are states the function of which is to keep their positive counterparts from being accepted.

Conditional opinions—opinions to the effect that if something is so-and-so then something is such-and-such—are states the function of which is to make someone accept that something is such-and-such given that one accepts that something is so-and-so.

Conjunctive opinions—opinions to the effect that something is so-and-so and something is such-and-such—are states the function of which is to keep otherwise separate opinions available for inference...¹⁰⁸

¹⁰⁷ Ibid., 88. I have replaced “optation” with “desire.” See note 99.

¹⁰⁸ Ibid., 88.

Björnsson then appeals to these analyses to give his analyses of logical intuitions. Consider the intuition that (D) logically follows from (A) and (B) and the intuition that (1) is inconsistent. Björnsson suggests that to have the former intuition is to “appreciate the functional connection between accepting the conjunction of (A) and (B) and accepting (D)” —where the functional connection is just that to accept the conjunction of (A) and (B) is to be in a state the function of which is to get one to accept (D).¹⁰⁹ The latter intuition, the intuition that (1) is inconsistent, is, according to Björnsson, “an appreciation of the functional conflict within the set which is independent of whether or not we accept any further opinions” —where the conflict is that to accept the conjunction of (A) and (B) is to be in a state the function of which is to *get one* to accept (D), while to accept (C) is to be in a state the function of which is to *keep one* from accepting (D). Björnsson suggests that “the [teleological] function of this appreciation is to keep us from accepting the conjunction [of (A), (B), and (C)], thus eventually allowing us a determinate attitude towards (D) or to direct our minds at more fruitful tasks.”¹¹⁰

Notice that the above analyses of negative, conditional, and conjunctive opinions are *general* along two dimensions. First, they are general in that they purport to cover both complex opinions with exclusively non-moral components—such as (A) and (C)—*and* complex opinions that have moral components—such as (A’) and (C’). This allows Björnsson to generalize his remarks about the logical intuitions concerning (1) to cover not only other non-moral cases but moral cases as well. For example, he can give exactly parallel analyses of the intuitions that (2) is inconsistent and that (D’) logically follows from (A’) and (B’). Second, the analyses of complex opinions are also general in the

¹⁰⁹ Ibid., 89. I’ve changed “(A*)”, “(B*)”, and “(D*)” to “(A)”, “(B)”, and “(D)”, respectively. See note 101.

¹¹⁰ Ibid.

sense that “opinion” is intended to cover both beliefs and desires. The emotivist can thus accept those analyses. This, in turn, allows the emotivist to accept the analyses of logical intuitions in terms of “appreciations” of functional connections and conflicts.

The first part of Björnsson’s strategy is now complete: we have general accounts of the roles that logical intuitions and complex opinions play in our thinking.¹¹¹ The question now is “whether and why we should expect to have such states given that our simple moral opinions are beliefs and desires, respectively.”¹¹² Let’s focus on intuitions of inconsistency. Assuming Björnsson’s analysis, why would such a state be “chosen by natural selection and our learning mechanisms”? How, in other words, do we (more precisely: how did our ancestors) benefit from the existence of intuitions of inconsistency? We have already seen Björnsson’s answer: these states benefit us by allowing us to either have a “determinate attitude” toward something or “to direct our minds towards more fruitful tasks.”

The expression “determinate attitude” is obviously important here. Björnsson has in mind states that we “express or act from.”¹¹³ More precisely, by “determinate attitude” he means a state that is either “included in the arsenal of inner maps that we use for orientation” or a desire or plan of action that we “take seriously.”¹¹⁴ I will now try to clarify this notion a bit.¹¹⁵

The contrast between determinate and non-determinate attitudes is supposed to be between cognitive and practical representations that govern behavior and those that do

¹¹¹ Ibid., 90.

¹¹² I’ve replaced “optations” with “desires”. See note 99.

¹¹³ Ibid., 93.

¹¹⁴ Ibid.; see also *ibid.*, 90-93.

¹¹⁵ Ibid., 90-93.

not. Consider a modified example of Björnsson's.¹¹⁶ A stick partially submerged in water looks bent but feels straight. There is no inconsistency here. However, when the time comes for action a decision has to be made. My behavior obviously cannot be governed at once *both* by the representation of the stick as straight *and* the representation of it as bent. At the end of the day, only one of these representations can be "included in the arsenal of inner maps that ... [I] use for orientation." The upshot of the decision about which of these representations to be governed by (Björnsson calls the decision an "act of judgment") is my determinate cognitive attitude vis-à-vis the shape of the stick. Suppose that my determinate attitude is the representation of the stick as straight. The representation of the stick as bent is then a cognitive representation that is *not* a determinate attitude. Parallel remarks can be made about practical representations. When I walk past the cookie jar I desire to eat a cookie. At the same time I desire to stick to my diet. Again, there is no inconsistency.¹¹⁷ But these desires cannot both be realized. So, when the time comes for action a decision ("act of judgment") again has to be made. This time the decision is about which desire will govern my behavior, which one I will "take seriously." The upshot of this decision is my determinate practical attitude vis-à-vis my eating a cookie. Suppose that my determinate attitude is the desire to stick to my diet. The desire to eat the cookie is then a practical representation that is *not* a determinate attitude.

Björnsson's key idea, as I understand it, is simple. We need determinate attitudes in order to survive.¹¹⁸ Hence intuitions of inconsistency benefit us by helping us to avoid entering states of mind in which we are hopelessly conflicted about what determinate

¹¹⁶ Ibid., 90-91.

¹¹⁷ At least there is no *logical* inconsistency. The example is from G.F. Schueler. "Modus Ponens and Moral Realism", *Ethics* 98 (April 1988): 500. Björnsson discusses the example to, among other things, make the point I am making in the body of the text. Björnsson, "Emotivists," 91.

¹¹⁸ Ibid., 91.

attitude to have.¹¹⁹ To be sure, there are times when we are better off not having any determinate attitude at all toward some matter. But intuitions of inconsistency benefit us in these circumstances as well, for the functional conflicts they help us avoid do not present indifference as an option. Someone who accepts (A), (B), and (C), for example, cannot happily lack a determinate attitude on the issue of whether telling one's little brother to lie makes one nervous. Furthermore, we need determinate *practical* attitudes no less than we need determinate *cognitive* attitudes.¹²⁰ That is, our need for desires and plans of action that we "take seriously" is just as great as our need for "inner maps that we use for orientation."

This completes the second part of the two-part strategy Björnsson outlines in the passage quoted earlier as that strategy applies to intuitions of inconsistency. A complete discussion of Björnsson's solution to the problem of reasoning would also address his answers to the "whether and why we should expect to have such states" question as it applies to intuitions of consequence and to negative, conditional, and conjunctive opinions. However, we are already in a position to see how Björnsson responds to the three worries concerning the emotivist's ability to explain the "important fact."

Björnsson's response to the third worry is obvious. The worry, recall, was that it is unclear whether complex opinions such as conditionals and negations that have moral opinions as components even make sense if emotivism is true and moral opinions are desires. Björnsson responds by giving accounts of the roles that complex opinions play in our overall thinking that are general along the second dimension of generality

¹¹⁹ See 91-93, 103.

¹²⁰ Ibid., 91-93, 103.

distinguished above. That is, he gives analyses of complex opinions that are as amenable to emotivism as they are to cognitivism.

I will discuss how Björnsson addresses the first and second worries as they apply specifically to intuitions of inconsistency; the extensions to intuitions of consequence should be obvious. Applied specifically to inconsistency, the first worry was that emotivists cannot give the most initially attractive explanation of why moral opinions seem to enter into relations of inconsistency: moral opinions seem to enter into relations of inconsistency because they really do enter into them, where ‘inconsistency’ is understood in its standard truth-theoretical sense. Björnsson’s response, as I understand it, is that while the emotivist cannot give *that* explanation, she *can* tell a story that is equally plausible: moral opinions seem to enter into relations of inconsistency because moral opinions can enter into functional conflicts that get in the way of our forming determinate attitudes or turning our minds to more fruitful tasks.¹²¹

The second worry applied specifically to intuitions of inconsistency was that emotivism seems to require two separate explanations for the fact that moral opinions seem to enter into relations of inconsistency and the corresponding fact that non-moral opinions seem to enter into such relations, as well as two separate explanations for the fact that we are averse to (what we naturally think of as) inconsistencies involving moral

¹²¹ This is my best guess. Björnsson is not very clear on how exactly he responds to the first worry. But note the following two quotes: “Notice that, although sadly neglected by much writing on logic and emotivism, the parallel treatments of cognitivist and non-cognitivist accounts is absolutely crucial to the issue at hand.” Ibid., 90. “What we have seen is how emotivism (and indeed cognitivism) can begin to explain our intuitions of inconsistency and consequence. To my mind, the emotivist sketch seems as promising as its cognitivist counterpart.” Ibid., 103. Unfortunately, it is unclear just what Björnsson takes the “cognitivist sketch” to be. However, there is some textual support for the idea that he identifies it with what I called the “most initially attractive explanation.” See *ibid.*, 86. That, together with the idea highlighted in the above two quotes that we should be concerned with determining the plausibility of the emotivist’s explanation of intuitions of inconsistency *relative to the plausibility of the cognitivist’s explanation*, would lend some support to my interpretation of Björnsson’s response to the first worry.

opinions and the corresponding fact that we are averse to inconsistencies involving only non-moral opinions. Yet, what we want, presumably, are single general explanations—general in the first of the above two senses of covering both moral and non-moral cases. Björnsson provides a clear response to this worry. The thought is that the emotivist *can* give single general explanations.¹²² She can say that non-moral opinions and moral opinions both seem to enter into relations of inconsistency because both beliefs and desires can enter into functional conflicts that get in the way of our forming determinate attitudes or turning our minds toward more fruitful tasks. Similarly, the emotivist can say that we are averse both to inconsistencies involving only non-moral opinions and to (what we naturally think of as) inconsistencies involving moral opinions because our survival depends on our being able to form determinate (*cognitive and practical*) attitudes.

This completes my presentation of Björnsson’s solution to the problem of reasoning, his attempt to show that the emotivist can adequately explain the “important fact” despite the three worries we have discussed. I will now identify two problems with the solution.

Two Problems with Björnsson’s Solution

The first problem is straightforward: if we accept Björnsson’s analysis of negative opinions, then we have to reject his analysis of intuitions of inconsistency. Let me explain. According to Björnsson, intuitions of consequence are “appreciations” of functional connections, and intuitions of inconsistency are “appreciations” of functional conflicts. What exactly do these two claims mean? What’s the cash-value of

¹²² See *ibid.*, 87

“appreciation”? Björnsson never tells us. That doesn’t really matter, however, since he ought anyway to be able to restate the above two claims in the same form in which he gives his analyses of complex opinions. How might this go for intuitions of inconsistency? Recall that Björnsson suggests that “the function of ... [the “appreciation” of the conflict in (1)] is to keep us from accepting the conjunction [of (A), (B), and (C)].” It seems, then, that Björnsson is proposing the following analysis of intuitions of inconsistency:

Intuitions of inconsistency—the sort of intuition we have about (1) and (2)—are states the function of which is to keep certain conjunctive opinions from being accepted.

But now we have a problem: Björnsson has reduced intuitions of inconsistency to a species of negative opinion, where a “negative opinion” is, as Björnsson puts it, an opinion “to the effect that something isn’t so-and-so.”¹²³ This can’t be right. Whatever an intuition of inconsistency is, it is not just an opinion to the effect that a certain conjunction *isn’t* the case. So, the above characterization of intuitions of inconsistency will not do, not at least if we accept Björnsson’s analysis of negative opinions.

Björnsson needs some way of capturing the modal element that is present in the thought that (1) (for example) is inconsistent. Perhaps he would respond by saying something like this: “Intuitions of inconsistency are distinctive in that they are states the function of which is to keep certain (conjunctive) opinions from being accepted *no matter what*.” The rough idea would then be that intuitions of inconsistency differ from negative opinions in that, although both have the role of keeping certain opinions from being

¹²³ Björnsson seems to be aware of this. When he sums up the emotivist’s explanation for the existence of logical intuitions and logically complex opinions, intuitions of inconsistency are conspicuously absent. *Ibid.*, 103. Presumably he thinks that they are covered by his remarks on negative opinions. See also the bottom of *ibid.*, 93 and the top of *ibid.*, 94. It is unclear whether Björnsson is also aware of the problem I am about to raise in the body of the text.

accepted, intuitions of inconsistency accomplish this by presenting the “poison marked” (conjunctive) opinion (to borrow Björnsson’s nice expression) as an *absolute non-option*, while negative opinions present the “poison marked” opinion as merely something to be avoided.¹²⁴ The difference is, as it were, in the strength of the poison. I will remain neutral on whether this rough idea can be turned into a satisfactory solution to the problem I have raised.

The second problem concerns Björnsson’s attempt to discharge the second part of his strategy as it applies to intuitions of inconsistency, that is, it concerns his answer to the question of “whether and why we should expect to have ... [states answering to Björnsson’s analysis of intuitions of inconsistency] given that our simple moral opinions are beliefs and desires, respectively.”¹²⁵ Obviously, for our ancestors to have benefited from the existence of a state the function of which is to keep one from accepting conjunctions like the conjunction of (A), (B), and (C) and the conjunction of (A’), (B’), (C’) it must have been *possible* for them to accept such conjunctions in the first place. But now consider the following remark Björnsson makes when addressing the “whether and why” question as it applies to conditional opinions:

Thinking that p might lead me to think that q , but I might already be thinking that not- q , or might perhaps be led from thinking that p to think that r which leads me to think that not- q : neither of these connections are my opinions that if p then q until judgment has determined which to follow
...¹²⁶

Björnsson is claiming here that one cannot have the opinion that if p then q at the same time one is both thinking that p and thinking that not- q . Now, it is unclear how we should read the expressions “thinking that p ” and “thinking that not- q .” Is thinking that p the

¹²⁴ Ibid., 94.

¹²⁵ I’ve replaced “optations” with “desires”. See note 99.

¹²⁶ Ibid., 98-99.

same thing as accepting the opinion that p ? If not, then what is the difference? The expression “are my opinions that if p then q ” also needs clarification. Is *having* an opinion the same thing as *accepting* an opinion? And how exactly do all of these states relate to the notion of a *determinate opinion*? I won’t attempt the difficult task of answering these questions but will instead make two assumptions. The first assumption is that one accepts an opinion only when one has that opinion, in whatever sense of the latter that is operating in “are my opinions that if p then q ”. The second assumption is that one has the opinion that p only if one is thinking that p , in whatever sense of “thinking” Björnsson has in mind. Once these assumptions are in place a problem emerges. According to Björnsson’s analysis, the role of intuitions of inconsistency is to keep us from accepting conjunctions like the conjunction of (A), (B), and (C), and the conjunction of (A’), (B’), and (C’).¹²⁷ But Björnsson’s remarks in the passage just quoted (together with my two assumptions) imply that *it is impossible for someone to accept such a conjunction anyway*. It would then be a mystery why intuitions of inconsistency would be “chosen by natural selection and our learning mechanisms.”

The upshot is that Björnsson’s solution fails to answer the “whether and why” question as it applies to intuitions of inconsistency even when we consider only paradigmatic beliefs. Björnsson’s solution to the problem of reasoning thus looks to be a failure by Björnsson’s own lights. I can, however, think of two possible responses to the objection I have just made. First, Björnsson can retract his claim that it is impossible for someone to have the opinion that if p then q at the same time that one is both thinking that p and thinking that not- q . He can then say that that claim was never a central part of his

¹²⁷ Let us assume, for the sake of argument, that Björnsson has found a way to avoid conflating intuitions of inconsistency with negative opinions. Let’s assume, in other words, that he has solved the first problem I raised.

solution anyway and conclude that my objection is a fallacious *ad hominem*.

Alternatively, Björnsson can argue that one or both of my two assumptions is unwarranted. The difficulty with both of these responses is that (a) they depend on the details of what Björnsson has in mind with the expressions “accepting the opinion that *p*,” “*p* is my opinion,” and “thinking that *p*,” and (b) it is far from clear what those details are. The result is that we are not in a position to fully evaluate either response. My objection can thus be understood provisionally: given that Björnsson does not have a certain story to tell about what is involved in accepting the opinion that *p*, *p*’s being one’s opinion, and one’s thinking that *p*, his solution fails to adequately answer the question of “whether and why we should expect to have ... [intuitions of inconsistency, understood in accordance with Björnsson’s analysis,] given that our simple moral opinions are beliefs...” However, I think there is reason to be skeptical that Björnsson can tell such a story, since once one gives the kinds of analyses of logically complex opinions that Björnsson gives, it is hard to see what work could remain for intuitions of inconsistency to do. For example, consider again (1). Given Björnsson’s analyses of conditional, conjunctive, and negative opinions, one can accept every member of (1) only if (A), the conjunction of (A) and (B), or (C) is failing to perform its function. So there seems to be no need for some *other* state that has the function of keeping one from accepting every member of sets like (1).¹²⁸

Conclusion

We have seen how Björnsson attempts to answer the three worries raised by the “important fact”—raised, in his words, by the fact that we “give up moral opinions to avoid what we naturally think of as inconsistencies, and ... come to have new moral

¹²⁸ Thanks to Brian Kierland for helping me to get clearer on this point.

opinions as a result of having other opinions from which the new ones seem to follow.”¹²⁹

At the center of this attempt are general analyses of logical intuitions and complex opinions based on the roles these states play in our thinking. By Björnsson’s own lights, the success of his solution depends crucially on the plausibility of those analyses and on the plausibility of the idea that nature and our learning mechanisms chose states with those roles. Yet, as it stands, Björnsson’s analysis of intuitions of inconsistency is manifestly implausible, since it implies that to think that a set of opinions is inconsistent is just to think that a certain conjunction is not the case. Furthermore, we have just seen how part of Björnsson’s account of the interplay between conditional and negative opinions seems to render certain paradigmatic intuitions of inconsistency otiose. Now, it may be that Björnsson has the resources to solve these two problems—though I have given a reason for thinking that at least the second problem is insoluble. But so long as they remain unsolved, Björnsson has not solved the problem of reasoning.

¹²⁹ Ibid., 85.

Works Cited

- Ayer, A.J. *Language, Truth, and Logic*, 2d ed. N.p.: Gollantz, 1946. Reprint, New York: Dover Publications, 1952.
- Björnsson, G. "Why Emotivists Love Inconsistency." *Philosophical Studies* 104 (2001): 81-108.
- Blackburn, Simon. "Attitudes and Contents." *Ethics* 98 (April 1988): 501-517.
- _____. *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press, 1998.
- _____. Blackburn, Simon. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Oxford University Press, 1984.
- Brink, David. *Moral Realism and the Foundation of Ethics*. Cambridge, England: Cambridge University Press, 1989.
- Dancy, Jonathan. *Moral Reasons*. Oxford: Blackwell Publishers, 1993.
- Dreier, James. "Expressivist Embeddings and Minimalist Truth." *Philosophical Studies* 83 (1996): 29-51.
- _____. "Internalism and Speaker Relativism." *Ethics* 101 (October 1990): 6-26
- _____. "Troubling Developments in Metaethics." *Nous* 36 (2002): 152-168.
- Geach, Peter. "Assertion." *The Philosophical Review* (October 1965): 449-465.
- Gibbard, Allan. *Thinking How to Live*. Cambridge, Massachusetts: Harvard University Press, 2003.
- _____. *Wise Choices, Apt Feelings*. Cambridge, Massachusetts: Harvard University Press, 1990.
- Hale, Bob. "Can There Be a Logic of Attitudes?" In *Reality, Representation, and Projection*, ed. John Haldane and Crispin Wright, 337-363. New York: Oxford University Press, 1993.
- Hurley, Susan. "Frege, the Proliferation of Force, and Non-cognitivism." *Mind* 93 (October 1984): 570-576.

- Jackson, Frank. "Non-cognitivism, Validity, and Conditionals." In *Singer and his Critics*, ed. Dale Jamieson, 18-37. Oxford: Blackwell Publishers, 1999.
- Joyce, Richard. "Expressivism and Motivation Internalism." *Analysis* 62 (October 2002): 336-344.
- Kölbel, Max. *Truth Without Objectivity*. London: Routledge, 2002.
- Lillehammer, Hallvard. "Smith on Moral Fetishism." *Analysis* 57 (July 1997): 187-195
- Miller, Alexander. *An Introduction to Contemporary Metaethics*. Cambridge, England: Polity Press, 2003.
- Ridge, Michael. "Sincerity and Expressivism." *Philosophical Studies* 131 (November 2006): 487-510.
- Schueler, G.F. "Modus Ponens and Moral Realism." *Ethics* 98 (April 1988): 492-500.
- Schroeder, Mark. "Expression for Expressivists." Forthcoming in *Philosophy and Phenomenological Research*.
- Shafer-Landau, Russ. *Moral Realism: A Defence*. Oxford: Oxford University Press, 2003.
- Searle, John. *Speech Acts: An Essay in the Philosophy of Language*. New York: Cambridge University Press, 1969.
- Smith, Michael. *The Moral Problem*. Oxford: Blackwell Publishers, 1994.
- Stevenson, Charles. "The Emotive Meaning of Ethical Terms." In *Moral Discourse and Practice: Some Philosophical Approaches*, ed. Stephen Darwall, Allan Gibbard, and Peter Railton, 71-82. Oxford: Oxford University Press, 1997.
- Stocker, Michael. "Desiring the Bad: An Essay in Moral Psychology." *The Journal of Philosophy* 76 (December 1979): 738-753.
- Stoljar, Daniel. "Emotivism and Truth Conditions." *Philosophical Studies* 70 (April 1993): 81-101.
- Svavarsdóttir, Sigrún. "Moral Cognitivism and Motivation." *The Philosophical Review* 108 (April 1999): 161-219.
- Van Roojen, Mark. "Expressivism, Supervenience, and Logic." *Ratio* 18 (2005): 190-205.
- _____. "Humeanism and Anti-Humeanism about Moral Judgments." *Philosophy and*

Phenomenological Research 65 (July 2002): 26-49.

_____. "Moral Cognitivism vs. Non-cognitivism." In *The Stanford Encyclopedia of Philosophy*, Winter 2005 ed., ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/win2005/entries/moral-cognitivism/>.

VITA

Jason Berntsen was born September 24, 1972 in The Bronx, New York, to Kenny and Mary Jane Berntsen and grew up in Houston, Texas. In 1995, he graduated from The University of Texas at Austin with a B.A. in philosophy. From 1995 to 1998, Jason taught at a Montessori pre-school in Austin. In 1998, Jason started the graduate program in philosophy at the University of Missouri-Columbia. He earned his M.A. in 2001 and his Ph.D. in 2007. Jason is currently an Assistant Professor of Philosophy at Xavier University of Louisiana.