

Debunking Challenges to Moral Realism

by

Matthew C. Braddock

Department of Philosophy
Duke University

Date: _____

Approved:

Walter Sinnott-Armstrong, Co-Supervisor

David Wong, Co-Supervisor

Owen Flanagan

Geoffrey Sayre-McCord

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Philosophy in the Graduate School
of Duke University

2012

ABSTRACT

Debunking Challenges to Moral Realism

by

Matthew C. Braddock

Department of Philosophy
Duke University

Date: _____

Approved:

Walter Sinnott-Armstrong, Co-Supervisor

David Wong, Co-Supervisor

Owen Flanagan

Geoffrey Sayre-McCord

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Philosophy in the Graduate School
of Duke University

2012

Copyright by
Matthew C. Braddock
2012

Abstract

Heightened awareness of the evolutionary, socio-cultural, and psychological origins of our moral judgments pushes many of us in the direction of moral skepticism, in the direction of doubting the objective truth of our moral judgments. But should awareness of the origins of our moral judgments shake our confidence in them? Are there good moral debunking challenges or debunking arguments from premises concerning the accessible origins of our moral judgments to skeptical conclusions regarding them? In vigorous pursuit of these questions, this dissertation sifts three promising moral debunking challenges to moral realism, namely Richard Joyce's (2001) evolutionary debunking argument from epistemic insensitivity, Sharon Street's (2006) "Darwinian Dilemma," and David Enoch's (2010) "Epistemological Challenge." It is argued that each challenge faces cogent objections that not only demonstrate the inadequacy of the best debunking challenges available but also instructively guide us to the development of new and more forceful debunking challenges to moral realism. This dissertation develops two new and forceful debunking challenges, both of which target the epistemic reliability and justification of our moral judgments on realist views of the moral facts. The first new debunking challenge starts from the premise that the best explanation of our moral judgments does not appeal to their truth and invokes a new species of epistemic insensitivity to secure the conclusion that our moral belief-forming

processes are epistemically unreliable. The second new debunking challenge reasons that the best explanation of the fact that moral realists have no good explanation of the reliability of our moral belief-forming processes is that there is no such reliability.

Contents

Abstract	iv
Acknowledgments	ix
Chapter 1: Introduction.....	1
Chapter 2: Moral Debunking Challenges from Epistemic Insensitivity	9
2.1 Introduction.....	9
2.2 Joyce’s Debunking Argument from Insensitivity	17
2.3 An Objection from Modal Distance	27
2.4 Charting Debunking Arguments from Insensitivity	34
2.5 A New Moral Debunking Argument from Insensitivity	37
2.6 Three Objections	47
2.7 Concluding Remarks.....	57
Chapter 3: Street’s Darwinian Dilemma and Enoch’s Epistemological Challenge	59
3.1 Introduction.....	59
3.2 Enoch’s Epistemological Challenge	61
3.3 Objections to Enoch’s Inference from No Good Explanation	78
3.4 Street’s Darwinian Dilemma.....	84
3.5 Objections to Street’s Inference from No Good Explanation	103
3.6 Concluding Remarks.....	106
Chapter 4: No Good Explanation of Moral Reliability	108
4.1 Introduction.....	108

4.2 Enoch’s Explanation.....	110
4.3 Copp’s Explanation.....	118
4.4 Clarifying the Targeted Shared Premise: Evolutionary Reliability	123
4.5 Assessing Evolutionary Reliability on Enoch’s and Copp’s Explanations.....	136
4.6 Assessing Evolutionary Reliability on Rationalist Explanations.....	149
4.7 Assessing Evolutionary Reliability More Generally: Arguments from Evolutionary Biases and Moral Nativism	166
4.8 Concluding Remarks.....	193
Chapter 5: A New Explanatory Challenge.....	196
5.1 Introduction.....	196
5.2 Target: Moral Realism.....	196
5.3 The Explanatory Challenge.....	198
5.4 Developing the Explanatory Challenge	205
5.4.1 <i>Gist of the Challenge</i>	205
5.4.2 <i>Against Causal Ignorance, Moral Ignorance, and Insufficient Effort</i>	209
5.4.3 <i>Against Cognitive Closure and Evolutionary Handicap</i>	214
5.4.4 <i>Against Bruteness and Chance</i>	220
5.4.5 <i>The Best Explanation: Eliminativism</i>	222
5.4.6 <i>Inference to the Best Explanation: What Follows?</i>	228
5.4.7 <i>Reaching the Skeptical Conclusion</i>	234
5.5 Objection: What Do Empirical Moral Psychologists Know?.....	239
5.5.1 <i>The Objection</i>	239

5.5.2 <i>Assessing the Objection</i>	240
5.5.3 <i>A Local Version of the Explanatory Challenge</i>	246
5.6 Concluding Remarks.....	254
Chapter 6: Conclusions and Directions for Future Research.....	255
References	258
Biography	265

Acknowledgments

My favorite part of a good scholarly book is usually the Preface, since often it provides a brief window into the humanity of the writer, a glance at the entourage of people surrounding the creation of the book in hand. After looking at many of these prefaces and writing a dissertation myself, I am convinced that good colleagues, friends, and family are necessary conditions for a scholarly book to come about, at least if it's any good. My good colleagues include my fantastic dissertation co-supervisors Walter Sinnott-Armstrong and David Wong, who lavished detailed and sharp written comments on my work again and again, often going well beyond their professional duties. Each interaction with them was of timely benefit to my dissertation and also a formative experience in philosophical training: as good mentors, they showed me how good philosophical reasoning and writing gets done. Looking over the dissertation, I can detect their fingerprints on every page. Their comments, questions, and objections often sent me in new and fruitful directions. Handsome thanks goes to them.

Kind thanks also to Owen Flanagan and Alex Rosenberg whose useful and encouraging comments during the early stages of my work gave me direction and confidence in my project. Thanks as well to Geoff Sayre-McCord who generously commented on my work and also, along with Owen, Walter, and David, raised excellent

challenges at my dissertation defense that have given me directions for further research. Given their expertise and willingness to help me, it is no exaggeration to say that my dissertation committee was the ideal committee for my project.

Many other colleagues also gave me useful feedback on earlier versions of the dissertation and content related to the dissertation. David Enoch kindly gave me insightful written comments and good conversation on earlier versions of Chapter 3 and 5. Kind thanks to Andreas Mogensen for our stimulating discussions during the Spring of 2012 when we co-taught a seminar at Duke on the content surrounding the dissertation: I learned a lot! Same goes to Daniel Korman, Sean D. Walsh, and Joshua Thurow for similarly stimulating discussions. Thanks also to participants of the Summer Seminar on Perceptual, Moral, and Religious Skepticism at Purdue University (2011) for their useful feedback on an earlier version of Chapter 5 of my dissertation and to Michael Bergmann, Pat Kain, and the Templeton Foundation for directing and sponsoring that excellent epistemology-packed seminar.

Much appreciation also goes to Justin Horn, Brian Besong, Kelly Becker, Tim Black, Steven Luper, Patrick Epley, and audiences at Central APA (2011), Society of Exact Philosophy (2011), Central States Philosophical Association (2011), and Univ. of Oklahoma (2011) for their refining comments on earlier versions of Chapter 2 of my dissertation. Same goes to Billy Dunaway and the audience at the Indiana Philosophical Association (2011) for comments on material related to Chapter 4 and 5.

Finally, many philosophers and empirical theorists who took the time to correspond with me about things like the nature of epistemic reliability, the mechanics of debunking challenges, the contingency of cultural evolution, and the nature of evolutionary biases. Among them, David Enoch, Kelly Becker, and Chandra Sripada particularly stand out for their useful and extended feedback in back-and-forth exchanges with me. Thanks to all of them.

The Philosophy Department and Graduate School at Duke University also deserve mention for providing an ideal environment in which to work: more than sufficient financial support from the Graduate School (e.g. Summer Research Fellowships, Bass Instructorship), collegial conversations and informal mentoring (e.g. Allen Buchanan, Andrew Janiak, Iris Einheuser), and always helpful and kind staff (e.g. Stacey Meyers, Janelle Haynes, Faisal Khan, and Karen Neander our Director of Graduate Studies).

My friends and officemates Gordon Steenbergen, Stephen Martin, and Nate Gindele also deserve my gratitude. Their encouragement, good humor, and camaraderie for the past couple years made writing this dissertation much more enjoyable than it otherwise would have been. Our stimulating conversations and our refreshing basketball games at noon on Tuesdays and Thursdays (or whenever we felt like playing) neatly broke up the day of research and writing for me. Core 4.

In most books the writer finishes the Preface, unsurprisingly, by thanking family. My case is no exception: my greatest and overwhelmingly large debt is to my family. My parents—Mark and Cindy—gave me unflagging support and gentle direction, as a youth and as a man. Our conversations together stick with me. Their encouraging love and the love and acceptance of my sisters, brother, grandparents, and aunt—Michele, Jessica, Stacy, Rachel, Shea, Luke, Nana Sue and Papa, Dawn and Jim, and Aunt Cathy—are sources of gladness and strength for me. Anne Marie’s side of the family—George and Becky, Susan and Ed, and Eric—also kindly gave me their friendship and financial support. Thanks to all my family for cheering me on.

Most importantly, my deepest gratitude is due to Anne Marie, my wife. Her unceasing support and companionship for the past seven years has been the best thing in the world. Through thick and thin, she has supported me. If any condition of my dissertation has been necessary, it has been Anne Marie’s love for me. We now have three (!) children and look forward to the road ahead of us.

Chapter 1: Introduction

Suppose we could rewind evolutionary history to the moment where our ancestors' moral judgments first began to emerge. Suppose we could then push the play button and visually observe their moral judgments unfold on the big screen, while we listen to running commentary by scientific experts. As the film progresses, we eventually get to observe the variegated origins and spread of our own moral judgments. After viewing the film, should we change the way we view our moral judgments? More specifically: should our confidence in the objective or response-independent truth of our moral judgments be shaken? Should we start doubting *all* our moral judgments or *some* of them (if so, which ones)? The answers to these questions appear to depend on what we observed, that is, on the accessible origins of our moral judgments.

Heightened awareness of the evolutionary, socio-cultural, and psychological origins of our moral judgments pushes many of us in the direction of moral skepticism, in the direction of doubting the objective or response-independent truth of our moral judgments. But should awareness of the origins of our moral judgments shake our confidence in them? Are there good moral debunking challenges or debunking arguments from premises concerning the accessible origins of our moral judgments to skeptical conclusions regarding them? In vigorous pursuit of these questions, this

dissertation sifts three promising moral debunking challenges to moral realism, namely Richard Joyce's (2001) evolutionary debunking argument from epistemic insensitivity, Sharon Street's (2006) "Darwinian Dilemma," and David Enoch's (2010) "Epistemological Challenge." It is argued that each challenge faces cogent objections that not only demonstrate the inadequacy of the best debunking challenges available but also instructively guide us to the development of new and more forceful debunking challenges to moral realism. This dissertation develops two new and forceful debunking challenges, both of which target the epistemic reliability and justification of our moral judgments on realist views of the moral facts, views which hold the moral facts are response-independent in the sense that they are not constituted by nor a function of (actual, ideal, individual, or collective) human responses and attitudes. The first new debunking challenge—call it *New Debunking Argument from Insensitivity*—starts from the premise that the best explanation of our moral judgments does not appeal to their truth and invokes a new species of epistemic insensitivity to secure the conclusion that our moral belief-forming processes are epistemically unreliable. The second new debunking challenge—call it *Explanatory Challenge*—reasons that the best explanation of the fact that moral realists have no good explanation of the reliability of our moral belief-forming processes is that there is no such reliability.

After the present introductory chapter, the second chapter – “Moral Debunking Challenges from Epistemic Insensitivity” – explores a classic debunking challenge to moral realism. It possesses this broad form: the best explanation of our moral judgments does not appeal to their truth (on moral realism), hence we are unjustified in holding our moral judgments. Recent discussion of this sort of argument fixates on the merit of the explanatory premise. But even if we grant the explanatory premise, it is an opaque matter how to get from there to the debunking conclusion. The apparent impotence of the explanatory premise is puzzling, since many have taken this premise (if true) to be plausibly pregnant with skeptical implications and thus worth refuting. Others have contended that the explanatory premise has no significant skeptical payoff. This chapter shows that the explanatory premise is indeed philosophically pregnant with skeptical implications by showing how to get from it to the debunking conclusion in a series of plausible steps involving a new species of epistemic insensitivity. The upshot is that we should really care about the merit of moral explanations, for otherwise moral skepticism looms. Rather than arguing in a dialectical vacuum, the chapter first engages a promising argument from the explanatory premise to the debunking argument, namely Richard Joyce’s (2001) evolutionary debunking argument from epistemic insensitivity. After reconstructing Joyce’s argument, I develop a couple of strong objections to it that instructively guide us to a new and better argument from

insensitivity, and then develop and defend this new argument from insensitivity which secures the inference from the explanatory premise to the debunking conclusion.

The third chapter – “Street’s Darwinian Dilemma and Enoch’s Epistemological Challenge” – inspects prominent ways of fleshing out a provocative and relatively neglected sort of debunking challenge, namely an explanatory challenge to moral realism that runs from considerations concerning the origins of our moral beliefs. Put broadly, the challenge is this: provide a plausible explanation of moral reliability or face moral skepticism. The challenge is motivated by the difficulty of explaining why we should rely upon or trust the evolutionary, socio-cultural, and psychological processes behind our moral judgments to lead us to objective (response-independent) moral truth. The explanatory challenge can morph into different forms with different scopes. Giving recent and rare attention to the challenge, Sharon Street (2006) and David Enoch (2010) develop sophisticated versions of it that possess different scopes. Enoch targets a robust or platonic form of moral realism with his Epistemological Challenge, while Street with her Darwinian Dilemma targets robust and non-robust versions of moral realism that accept the response-independence of the moral facts. This chapter reconstructs both of their challenges and puts a finger on an important and unnoticed central defect that they both share, namely their inability to secure the inference to skepticism about moral reliability from the claim that we have no plausible explanation of that reliability. This

objection not only shows that the two most prominent explanatory challenges in this genre do not pass muster, but it instructively guides us to the construction of a new and more forceful explanatory challenge to moral realism, which is developed and backed up in the next two chapters.

The fourth chapter—“No Good Explanation of Moral Reliability”—argues for a central premise that figures in explanatory challenges like Street’s Darwinian Dilemma, Enoch’s Epistemological Challenge, and the new explanatory challenge to be developed in the next chapter. The premise is that moral realists possess no good explanation of the reliability of the processes producing our moral judgments. This premise deserves backing because moral realists, in response to the explanatory challenges of Street (2006) and Enoch (2010), have developed sophisticated explanations of moral reliability designed to cohere with our best empirical understanding of the origins of our moral judgments. Accordingly, this chapter evaluates two prominent realist explanations—David Copp’s (2008) and David Enoch’s (2010)—and also a prominent and empirically-informed rationalist explanation developed by Peter Singer (1981, 1999) that could be invoked by realists to help explain moral reliability. I argue in an empirically-informed way that these prominent explanations are implausible, since they do not vindicate a shared epistemic claim that not only they rely upon but that any plausible realist

explanation of moral reliability apparently must rely upon, namely the claim that the evolutionary processes behind our initial¹ moral judgments are sufficiently reliable.

Though the empirically-informed critique of prominent realist explanations motivates the premise that moral realists possess no plausible explanation of moral reliability, additional support is desirable, since it is quite possible that alternative realist explanations, perhaps not yet conceived, could do better with the empirical and philosophical resources at hand. This chapter dims the latter prospect, however, by drawing from recent work in empirical moral psychology and the evolutionary sciences to develop two empirically-based arguments for thinking not only that Copp's, Enoch's, and Singer's respective explanations do not work, but that no plausible alternative realist explanation of moral reliability is currently available, given our best empirical understanding of the evolutionary biases and nativist psychology behind our moral judgments.

By shoring up the claim that moral realists have no plausible explanation of moral reliability, this chapter strengthens explanatory challenges like Street's, Enoch's, and the new one to-be-developed in the next chapter. This is interesting because a central critique of Street's Darwinian Dilemma—for example, developed by David Copp

¹ Let *initial moral judgments* refer to noninferentially formed judgments which are in turn fed into our reasoning processes.

(2008) and Clarke-Doane (2012)—is that she has not successfully shown that realists have no good explanation of moral reliability. Similarly, Enoch (2010) argues that the Epistemological Challenge he reconstructs is ultimately defective because there is in fact a plausible explanation available, namely Enoch’s own evolutionary explanation which he develops in the same paper. But if my fourth chapter successfully shows realists—including Copp (2008) and Enoch (2010)—possess no plausible explanation of moral reliability, then it neutralizes central objections to the challenges of Street and Enoch as well as to the new challenge developed in the next chapter.

The fifth chapter — “A New Explanatory Challenge” — develops a new explanatory challenge to moral realism, call it *Explanatory Challenge*. It challenges the moral realist to either provide a plausible explanation of the process reliability of our moral judgments that coheres with our best empirical understanding of the origins of those judgments or face moral skepticism. The challenge developed is based on an inference to the best explanation that reasonably secures (or better secures) the inference that Street’s (2006) Darwinian Dilemma and Enoch’s (2010) Epistemological Challenge could not secure (as shown in Chapter 3). The *explanandum* or thing to be explained is the moral realist’s failure to plausibly explain the reliability of our moral belief-forming processes. According to the challenge, the best explanation of the moral realist’s failure to explain moral reliability, on standard criteria of explanatory merit, is that there is no

such reliability. From this explanatory outcome, it is argued that skeptical implications reasonably follow for the justification of our moral judgments. The chapter also sketches an interesting local version of the challenge, which shows that even if the original (generalized) *Explanatory Challenge* were unsuccessful, a local version could still possess significant skeptical bite.

The sixth and final chapter—“Conclusion and Directions for Future Research”—briefly sums up major conclusions of the dissertation and sketches directions for further research.

Chapter 2: Moral Debunking Challenges from Epistemic Insensitivity

2.1 Introduction

A heightened awareness of the evolutionary, socio-cultural, and psychological origins of our moral judgments pushes many of us in the direction of moral skepticism, in the direction of thinking we are unjustified in holding our moral judgments on a realist or objectivist understanding of the moral facts as response-independent. Why? Should it? We need an argument. A classic argument that fleshes out this skeptical worry from origins moves from the best explanation of our moral judgments to their debunking.¹ Call it *Classic Moral Debunking Argument*:

Classic Moral Debunking Argument

Explanatory Premise: The best complete explanation of our moral judgments does not appeal to their truth.

¹ Discussions of this sort of moral debunking argument took off after Gil Harman (1977, Ch. 1) initially developed an ambitious version of it. He argued, roughly, that since the best explanation of our moral beliefs does not appeal to the facts that would make them true, we have no reason for thinking moral facts (if construed as irreducible) exist. His bold argument triggered an avalanche of literature on “moral explanations.” Recently, Richard Joyce (2001, Ch. 6; 2007, Ch. 6) has developed sophisticated versions of the moral debunking argument.

Debunking Conclusion: So, we are unjustified in holding our moral judgments.²

The lion's share of the discussion surrounding this sort of argument has fixated on the merit of the Explanatory Premise with respect to moral realism, that is, with respect to views on which the moral facts are not constituted by nor a function of human attitudes and responses.³ Nicholas Sturgeon (1984) and others have no truck with the Explanatory Premise and instead champion "moral explanations" — explanations that appeal to moral facts realistically construed — as part of the best complete explanation of our moral judgments. For example, some contend that the fact that it is wrong to douse

² The Debunking Conclusion is cast in terms of *doxastic* justification rather than *propositional* justification. That is, the conclusion is that though our moral judgments may very well be *propositionally* justified in the sense that there are good reasons to think they are true or correct, they are nonetheless *doxastically* unjustified because our moral judgments are not *based on* such reasons or that we do not hold them *because* of such reasons. Most philosophers take doxastic justification to be a necessary condition for *knowledge*, so if the Debunking Conclusion holds, it may also plausibly imply knowledge-skepticism. Of course some epistemic externalists buck tradition and hold that justification is *not* necessary for knowledge (e.g. Kornblith 2008), and if they are right, doxastic justification-skepticism would not imply knowledge-skepticism. Nonetheless, such anti-justificationist externalists should still care about the argument for the Debunking Conclusion to-be-developed in the present chapter, since it reaches the Debunking Conclusion by means of the premise that the processes producing our moral beliefs are *unreliable*. Since these externalists typically hold process reliability to be a necessary condition for knowledge, my argument would, if successful, imply knowledge-skepticism even on prominent anti-justificationist accounts of knowledge.

³ For a nice review of the explanationist literature, see Majors (2007).

the cat with gasoline and light it up for kicks partly best explains why we judge it wrong to do so.⁴ Gilbert Harman (1977) and others doubt that such realist moral explanations are part of the best explanations of our moral judgments and so find themselves on board with the Explanatory Premise.⁵

But what hinges on this explanatory debate? First look, the impetus to shoot down the Explanatory Premise appears misplaced, since, as Walter Sinnott-Armstrong points out, it is opaque how the Explanatory Premise could get us to any kind of interesting skeptical conclusion regarding moral justification, moral knowledge, or the existence of the moral facts. Yes, the Explanatory Premise (if true) would rule out *explanationist*⁶ sources of justification for our moral judgments, but it would do nothing to rule out coherentist, experiential,⁷ intuitionist, or reliabilist sources of justification.⁸ Or so it seems. So how, without independently shooting down coherentist, experiential,

⁴ Also see Railton (1985) for the case that some moral facts best explain some non-moral phenomena (e.g. moral facts regarding injustice partly best explain the social tendency to revolt).

⁵ Also see Leiter (2001), Sommers and Rosenberg (2003), Joyce (2001, Ch. 6; 2007, Ch. 6).

⁶ Explanationist routes to vindicating our moral judgments are arguments that our moral judgments are justified because the moral facts that make them correct play a role in their best explanation.

⁷ Let *experiential* sources of justification refer to moral experience, moral seemings, or their kin.

⁸ Sinnott-Armstrong (2006), p.44.

intuitionist, and reliabilist moral epistemologies, can we get from the Explanatory Premise to the Debunking Conclusion? One naturally draws a blank.⁹ The apparent impotence of the Explanatory Premise is puzzling since many have taken the Explanatory Premise to be pregnant with skeptical implications and thus worth refuting, that is, something that moral anti-skeptics should take seriously.¹⁰ If the Explanatory Premise has no skeptical force, however, this suggests it does not deserve the attention it has received and continues to receive.

Enter this chapter. This chapter shows that the Explanatory Premise is philosophically pregnant with skeptical implications by showing how to get from it to the Debunking Conclusion in a few plausible steps. By showing how to get from here to there, the chapter shows that moral anti-skeptics—that is, advocates of the view that we are justified in holding our moral judgments—must reject the Explanatory Premise and

⁹ Richard Joyce, in his later book *The Evolution of Morality* (2007, pp. 214-215), attempts rather quickly to infer unreliability from the Explanatory Premise by appeal to an analogy, but Wielenberg (2010, pp. 461-463) offers the decisive critique of the analogy. Nothing more needs to be added to what Wielenberg says. It is entirely unobvious how to move from the Explanatory Premise to unreliability, and Joyce does not show how. Joyce also quickly argues in the same section that the Explanatory Premise impugns coherentist and intuitionist sources of moral justification (see Joyce (2007), pp. 216-219)). Throughout this discussion Joyce appears to suppose it is obvious that the Explanatory Premise defeats whatever justification these sources might initially confer, but it is far from obvious to people like Walter Sinnott-Armstrong (see previous note) and the present author. Joyce needs bridge premises. The present chapter aims to supply them.

¹⁰ For a good discussion of the alleged skeptical implications of the Explanatory Premise, see Pust (2001).

accept at least some moral explanations as best in order to preserve their anti-skeptical stance. The fact that moral anti-skepticism hinges on the success of moral explanation is a significant outcome which brings a renewed sense of urgency to the debate over moral explanations. That is, we should really care about the success of moral explanations, for otherwise moral justification-skepticism looms.

Interestingly, it is often thought that some realist metaethical accounts have a harder time vindicating moral explanations than others. For example, robust ethical non-naturalists typically eschew advancing moral explanations, and not without reason: it is not easy to see how the (e.g. non-spatiotemporal, causally inert) non-natural facts they posit play a role in the best explanations of our moral judgments.¹¹ Realist versions of ethical naturalisms that construe moral facts as response-independent natural facts are not out of the woods either—they owe us a case for moral explanations as part of the best explanations of our moral judgments—but it is often thought that their explanatory job is not as difficult, at least in principle, since they can appeal to moral facts *qua* natural facts as playing a causal role in the best explanation of our moral judgments. If this dialectical asymmetry holds, then an argument that shows us how to get from the

¹¹ See Majors (2007), however, for the case that ethical non-naturalists should also advance moral explanations.

Explanatory Premise to the Debunking Conclusion would appear to hit hardest (though not exclusively) against ethical non-naturalisms.

Good bridge premises, of course, must do the work of getting us from the Explanatory Premise to the Debunking Conclusion. The hunt for good bridge premises in this chapter proceeds not in a dialectical vacuum but rather by way of inspecting what appear to be among the most promising bridge premises on offer, namely those offered by Richard Joyce (2001, Ch. 6) in his important but almost entirely neglected¹² evolutionary moral debunking argument from insensitivity. From an evolutionary version of the Explanatory Premise, Joyce infers the counterfactual claim that the evolutionary processes producing our moral judgments are *insensitive* in the sense that were there no moral facts whatsoever, the evolutionary processes producing our moral judgments would lead us to the very same moral judgments (i.e. judgments with the same content). The argument then takes the insensitivity of the evolutionary processes producing our moral judgments to imply the unreliability of such processes and then infers the Debunking Conclusion from unreliability.

Joyce's evolutionary moral debunking argument from insensitivity is important because though neglected it is among the best moral debunking arguments available.

¹² Though Joyce published this argument ten years ago, Sinnott-Armstrong (2006, p. 44) is the only one who gives it any sustained attention (and half a page at that). See Note 29 of the present chapter for a response to Sinnott-Armstrong's commentary on Joyce.

Moreover, it is important because, though subject to a couple of strong objections as it stands, its form inspires another argument from a new species of epistemic insensitivity that shows how to get from the Explanatory Premise to the Debunking Conclusion, as the present chapter will show. This chapter reconstructs Joyce's argument, develops a couple objections to it that point the way toward improvement, and, building upon this discussion, develops and defends a new and better argument from insensitivity which secures the inference from the Explanatory Premise to the Debunking Conclusion.

Consider a few more snappy thoughts that should whet appetites for this chapter. The chapter should prove interesting to moral philosophers because any ambitious argument for the Debunking Conclusion is interesting, especially an argument that could do better than other arguments from the Explanatory Premise. Furthermore, nearly everyone has reason to care about this dialectic because the Explanatory Premise is widely thought to apply to many (if not all) of our moral judgments. Many also suppose that the same sort of premise applies to many (if not all) philosophical (e.g. metaphysical)¹³ and religious¹⁴ judgments. If so, then the debunking argument from insensitivity developed in this chapter, if successful, may have

¹³ For discussion, see Goldman (1989), (1992).

¹⁴ For discussion, see Murray (2009), pp. 173-176.

surprising and important debunking implications for non-moral judgments that many hold dear.

The chapter should also prove interesting to mainstream epistemologists because discussions of epistemic insensitivity almost always center on whether some sort of Nozick-style (1981) sensitivity condition is necessary for *knowledge*—that is, whether for a belief to be known it must be sensitive in the sense of having the following modal feature: if *p* were false, *S* would not believe that *p*. Nozick's sensitivity condition has encountered strong counterexamples and objections.¹⁵ But if there is a new species of epistemic insensitivity that avoids such problems and plausibly has implications for the *process reliability* of our judgments, as I shall argue in this chapter, then epistemologists will be interested in this upshot, especially if this sort of insensitivity follows from an Explanatory Premise which is widely thought to apply to many judgments in many different domains.

¹⁵ Nozick's sensitivity condition still has some current defenders (e.g. Becker 2008) but is currently unpopular and has been subjected to a barrage of counterexamples—for example, cases of inductive knowledge (Sosa 1999), higher-level knowledge (Vogel 2000), the charge that it unacceptably violates closure under known implication. See Luper-Foy (1993) for a shower of such objections and more. Fortunately, I am not defending Nozickean sensitivity as a necessary condition of knowledge (or justification or reliability), so these objections can be set aside.

2.2 Joyce's Debunking Argument from Insensitivity

To understand Joyce's (2001) moral debunking argument from insensitivity¹⁶, we must first understand his specific version of the Explanatory Premise on which it relies. Joyce offers a nativist evolutionary explanation of our tendency to make the moral judgments we do. In a word, natural selection endowed us with an innate (genetically encoded) moral sense or conscience. In more words: helping behavior evolved because it tended to promote reproductive fitness and the "moralization" of these behaviors—that is, coming to employ "categorical" moral concepts such as *obligation* and *forbiddenness* and to believe that we have binding, authoritative moral obligations—tended to bolster motivation to perform them and thus tended to make us more reliable cooperators than we would otherwise be were we motivated purely by self-interested desires or sympathy. Given the individual and group fitness benefits resulting from reliable

¹⁶ Michael Ruse (1986, p. 254) deserves credit for offering the first notable though sketchy moral debunking argument from insensitivity. Ruse infers the insensitivity of our moral beliefs—more specifically, the modal fact that if there were no objective moral facts, we would still arrive at the same moral judgments—from his version of the Explanatory Premise, that is, from the claim that his Darwinian sentimentalist explanation of our moral judgments, as the complete best explanation, does not appeal to their truth. Then he infers the nonexistence of objective moral facts from the insensitivity of our moral beliefs. But there is a noticeable gap in the latter inference. The modal fact that if there were no moral facts, we would still believe there were does not imply that there are no moral facts. Ruse's argument from insensitivity fails because it attempts to prove too much. The insensitivity of our moral beliefs (on whatever conception of insensitivity we plug in) simply cannot show that there are no moral facts and thus cannot do any metaphysical debunking.

cooperation, natural selection provided us with an “innate” (genetically encoded) moral sense, or more precisely, the psychological dispositions and mechanisms that give rise to a tendency to form judgments employing categorical moral concepts, judgments which actually form given normal environmental cues. Our moral minds are thus not blank slates but consist of a constellation of adaptive dispositions that, given the relevant environmental input, drive us to make the categorical moral judgments that we do.¹⁷

Joyce sketches and supports this nativist evolutionary explanation of our moral judgments.¹⁸ However, it is important to notice, as Joyce does, that to explain our evolved *disposition* to make certain sorts of judgments is not to *completely* explain our particular judgments.¹⁹ Evolved dispositions by themselves cannot do all the causal work in determining the content of our moral judgments, for the environment must play its part. When we form many (if not all) of our moral judgments, we do so in response to environmental cues or certain features of our experience. Surprisingly though, Joyce says next to nothing about the sorts of environmental cues that activate our evolved

¹⁷ Joyce (2001), Ch. 6; also see Joyce (2007), Chs. 1-4.

¹⁸ Joyce (2001, Ch. 6) assigns the status of “plausible speculation” (p.135) to this hypothesis and backs it up mainly with references to the empirical literature. Joyce (2007, Chs. 1-4) presents much more detailed support for it.

¹⁹ Joyce (2001), p. 147.

dispositions, and one gets the impression from what he does say that since we should only expect *not* to see these dispositions activated in aberrant environments (e.g. “a child raised isolated in a windowless room”), normal human environments will typically contain the relevant activating cues.²⁰ As Joyce’s broad gesture toward the environment suggests, he apparently has offered only a plausible explanation of our *disposition* to make categorical moral judgments rather than a *complete* explanation of the particular judgments themselves. This will prove important, as we shall see later in this section.

Joyce hitches his nativist explanatory hypothesis to the metaethical premise that his nativist evolutionary explanation does not appeal to the truth of our categorical moral judgments. More precisely, he argues that his explanation does not posit any facts that could *plausibly* be the moral facts. He argues for this point by pressing objections to ethical naturalisms which initially might be thought to construe the moral facts in such a naturalistically respectable way that they could plausibly play a role in Joyce’s evolutionary explanation of our categorical moral judgments. Most prominently, Joyce presses the objection that ethical naturalisms cannot make sense of the intuitively categorical or binding character of the moral facts.²¹ Hence, through the combination of

²⁰ Joyce (2001), p. 147; Cf. p. 162.

²¹ Joyce (2001), pp.148-157.

this metaethical premise and the nativist explanatory premise, Joyce arrives at his version of the Explanatory Premise. For Joyce, this combination makes for a potent skeptical cocktail.

Joyce (2001) conveniently lays out his evolutionary moral debunking argument from insensitivity in the following central passage:

On the assumption that my favored hypothesis about the ‘moral sense’ is correct, it follows that the process by which humans form moral judgments is an unreliable one, for they are disposed to do so regardless of the evidence to which they are exposed. Suppose that the actual world contains real categorical requirements—the kind that would be necessary to render moral discourse true. In such a world humans will be disposed to make moral judgments (most generally, to believe that categorical requirements exist), for natural selection will make it so. Now imagine instead that the actual world contains no such requirements at all—nothing to make moral discourse true. In such a world humans will *still* be disposed to make these judgments...just as they did in the first world, for natural selection will make it so. What this shows is that the process that generates moral judgments exhibits an independence relation between judgment and truth, and these judgments are thus unjustified. (162-3)

For clarity and ease of reference, Joyce’s reasoning can be reconstructed as follows:

Joyce’s Debunking Argument from Insensitivity

- (1) Joyce's nativist evolutionary explanation is best: natural selection endowed us with an innate disposition to make moral judgments.²²
- (2) Joyce's nativist evolutionary explanation does not appeal to the truth of our moral judgments.
- (3) If the best explanation of our disposition to make moral judgments does not appeal to their truth, then the processes producing our moral judgments are insensitive to the truth of their content (i.e. were there no moral facts, the processes that produce our moral judgments would lead us to the same moral judgments).²³

²² Joyce specifically targets our *categorical* moral judgments, such as our judgments of what we are morally required or obligated to do, or what would be wrong or forbidden. He does not say anything about our judgments of moral value or virtue, that is, what we think we have non-categorical moral reason to do or what traits we think are good to have. It is important to note that this absence narrows the scope of Joyce's version of the debunking argument. To avoid the cumbersome expression "categorical moral judgments", I shall drop the "categorical" qualifier from here on in the chapter.

²³ More precisely, let "processes producing our moral judgments" refer to *instantiations of process types* producing our moral beliefs. The more precise but too cumbersome reading of premise (3) would run as follows: the process types whose instantiations produce our moral judgments are insensitive to the truth of their content in the sense that were there no moral facts in worlds as otherwise similar as possible to our own (where the belief-forming process types are held constant), instantiations of the process types whose actual instantiations produce our moral judgments would lead us to the same moral judgments (i.e. judgments with the same moral content) at which we arrive in the actual world.

- (4) So, the processes that produce our moral judgments are insensitive to the truth of their content. (1-3)
- (5) If the processes that produce our moral judgments are insensitive to the truth of their content, then they are unreliable.
- (6) So, the processes that produce our moral judgments are unreliable. (4, 5)
- (7) If the processes that produce our moral judgments are unreliable, then we are unjustified in holding our moral judgments.
- (8) So, we are unjustified in holding our moral judgments. (6, 7)

The reasoning runs this way. If Joyce's nativist evolutionary explanation of our moral judgments is on track, then since such an explanation makes no appeal to the truth of our judgments, the following modal fact holds which we may call the *insensitivity* of the processes producing our moral judgments: were there no moral facts, then naturally selected dispositions, as part of the developmental process leading to our judgments, would still drive us to believe there were. Natural selection, as the best explanation of our moral judgments, would "make it so." Joyce supposes this fact of

insensitivity to show (somehow) that our moral judgments are “unreliable” and thus unjustified.²⁴

My initial criticism of Joyce’s argument zeroes in on premise (3). The insensitivity of our moral judgments does not follow simply from Joyce’s explanation of our *disposition* to arrive at them. As a general matter, an explanation of *S*’s mere disposition to believe that *p* does not imply *S* would believe *p* if *p* were false, because were *p* false it is entirely possible that *S*’s disposition to believe *p* would not be activated by the relevant cues. If Joyce only explains our disposition to make categorical moral judgments, he has *a fortiori* only offered a *partial* explanation of our moral judgments, but in order to know that we would come to the same judgment even if there were no moral facts—that is, to know that Joycean insensitivity holds—we need to know that the *complete* causal process leading to our moral judgment does not appeal to the truth of such judgments. After all, if moral facts play a causal role in part of the complete causal process leading to the belief, then were we to eliminate such facts, we should have no confidence in thinking that we would come to the same moral judgments.

Could Joyce repair his argument? Yes. First, he could supplement his nativist evolutionary explanation in order to offer a *complete* causal explanation of our particular

²⁴ Joyce (2001, pp. 163-64) avoids the anticipated “companions in guilt” objection that his argument applies to our perceptual and mathematical judgments by maintaining that the best explanation of our *perceptual* and *mathematical* judgments appeals to their truth.

moral judgments. Second, he could argue that this supplementary explanation, like the nativist part of his explanation, does not appeal to the truth of our moral judgments.

What else then, in addition to our evolutionary dispositions, does Joyce need to completely explain our moral judgments?²⁵ Plausibly, some socio-cultural-psychological explanatory supplement, which would include the relevant environmental cues that activate the dispositions Joyce posits. If Joyce had this explanatory supplement in hand, he would be in a better position to infer insensitivity. But he would also need to effectively argue, and this is no easy task, that this new supplementary explanation also does not appeal to the truth of our moral judgments. Furthermore, he would need to argue that his explanation does not *imply* the truth of our moral judgments. This additional condition is required to secure the inference to insensitivity because even if the best explanation of our moral judgments does not *appeal* to the moral facts to explain

²⁵ One way Joyce could make the repair is simply to extend his nativist evolutionary explanation of our *disposition* to make moral judgments to all the actual particular judgments themselves, that is, to argue that *every* particular moral belief (e.g. slavery is wrong) we have is selected for or hardwired in such a way that the nativist process producing the belief does not appeal to its truth. But this sort of evolutionary explanation would be overly adaptationist. First, *plasticity* of our moral belief system, rather than fixation, would most likely be reproductively advantageous. Second, it is not at all clear that many, much less *all*, of our moral beliefs were produced *and* sustained because they were adaptively useful. Even though *some* of our moral beliefs may be adaptations (or by-products of psychological tendencies that are themselves adaptive), many of our moral beliefs notoriously appear more like products of maladaptive social construction, and many others appear (more sweetly) like products of good reasoning or “rational intuition.”

anything, it may nonetheless logically imply them through a supervenience relation holding between the moral facts and the natural facts that are appealed to in the best explanation of our moral judgments. So then, if Joyce could develop and support his explanatory supplement, and if he could reasonably make the case that it does not appeal to *nor imply* moral facts, then he could reasonably infer insensitivity, for if the moral facts make no difference whatsoever as to why we come to the moral judgments we do and do not supervene on natural facts that make such a difference, then in the nearest moral factless world we would still arrive at the same moral judgments.

If we take this critical discussion into account, Joyce then could repair his argument and secure the inference to insensitivity in the following way:

Joyce's Debunking Argument from Insensitivity (repaired)

- (1*) A complete nativist, socio-cultural, psychological explanation of our moral judgments is best: natural selection endowed us with an innate disposition to make moral judgment, and this disposition is activated by familiar socio-cultural-psychological cues.
- (2*) This best complete explanation does not appeal to *nor imply* the truth of our moral judgments.

(3*) If the best complete explanation of our moral judgments does not appeal to *nor imply* their truth, then the processes that produce our moral judgments are insensitive to the truth of their content (i.e. were there no moral facts, the processes that produce our moral judgments would lead us to the same moral judgments).

(4)-(8). (Same as before, see p. 22).

This repaired Joycean argument is worth taking seriously and stands as an improvement over Joyce's original argument because Joyce needs a more complete explanation for his inference to insensitivity to get off the ground and because the sorts of familiar causes of our moral judgments identified by social scientists and psychologists that would be part of the explanatory supplement do not obviously appeal to the moral facts.²⁶ After all, many are skeptical of moral explanations of our moral judgments. Premise (3*) is an improved premise and improved in a way that does not impair premises (1*) or (2*). Premises (1*) and (2*) are challengeable of course, but since my interest in this chapter is not in seeing whether the Joycean version or any other version of the Explanatory Premise can be vindicated, but rather in seeing whether,

²⁶ For argument along these lines, see Leiter (2001, p. 94).

granting the Explanatory Premise, we can secure the inference to the Debunking Conclusion, let us not dwell on the first two premises further. I cared to revise them to show how a *properly revised* version of Joyce's Explanatory Premise would imply insensitivity. Having secured the move from the Explanatory Premise to insensitivity, we can now turn our attention to the other crucial link in the chain.

So far, we have seen that the move from the Explanatory Premise to insensitivity works *if* we construe the premise as saying that the *complete* best explanation of our moral judgments does not appeal to *nor imply* their truth. The move from unreliability²⁷ to the Debunking Conclusion is difficult to contest—good evidence that a process behind our judgments is unreliable (i.e. unlikely to produce true judgments) provides us with good reason for thinking we are not justified in holding those judgments. But the move from insensitivity to unreliability is subject to formidable challenge, as we shall see in the next section.

2.3 An Objection from Modal Distance

Let me develop the main objection to Joyce's argument, which sticks even if we repair Joyce's argument in the way suggested in the previous section. The move from the (properly revised) Explanatory Premise to Joycean insensitivity works, for if indeed

²⁷ That is, accessible unreliability or evidence of unreliability given by attention to the debunking argument from insensitivity.

the moral facts make no difference whatsoever to why we come to our moral judgments, then were there no such moral facts, we would come to the same moral judgments. But the move from Joycean insensitivity to unreliability does not work, or so I'll argue.

Before developing this objection to the inference from insensitivity to unreliability, let me sketch Joyce's motivation for the inference, since in order to see whether that inference holds up we should attend to the support he offers for it. Joyce summarizes the argument briefly at the beginning of the following paragraph and then supports it by analogy to paranoid belief:

The innateness of moral judgments undermines these judgments being true for the simple reason that if we have evolved to make these judgments irrespective of their being true, then one could not hold that the judgments are *justified*. And if they are unjustified, then although they *could* be true, their truth is in doubt. Consider an analogy. John makes judgments that Sally is "out to get him." After talking to John's psychiatrist, we discover that John is neurotically paranoid about Sally, and would form these judgments about her regardless of what she did. If whatever evidence John is provided with concerning Sally's intentions, he will make exactly the same paranoid judgment, then there is absolutely no reason to think that his judgments are true. (2001, p. 159)

The idea is that just as the insensitivity of the process producing John's paranoid belief implies the unreliability and debunking of that belief, so the insensitivity of the processes producing our moral beliefs implies the debunking of these beliefs. So much for Joyce's motivation. Its assessment must wait until the end of this section, after we have laid the epistemological groundwork.

Now onto the main objection. Joyce asks us to consider a world that is just like the actual world except that there are no moral facts. In that world, Joyce contends, we would still arrive at the same moral judgments at which we arrive in the actual world and thus our moral judgments are unjustified because they are “unreliable.” A key interpretive question: what conception of epistemic reliability is Joyce operating with? He offers no explicit specification, so it is an open question. Theoretically, he has a number of options in philosophical space.²⁸ This textual evidence suggests that Joyce is most likely operating with a standard conception of process reliability. That would also make charitable sense of why he does not offer an explicit characterization of reliability.²⁹

²⁸ E.g. reliability as sensitivity, safety, Nozickean tracking, proper functioning, epistemically virtuous belief-formation, reliable indication, actual world process reliability, and counterfactual process reliability (“normal worlds” reliability (see Goldman (1986, p. 107)), actual-plus-nearby-worlds reliability).

²⁹ Sinnott-Armstrong (2006, p.44) sees Joyce as *conflating* actual-world process unreliability with the distant counterfactual process unreliability or insensitivity that Joyce invokes and contends that though the former matters, it is not at all clear that the latter matters. In contrast, I see Joyce not as *conflating* (or confusing or identifying) actual-world process unreliability with insensitivity but rather as *inferring* standard (actual-plus-nearby world) process unreliability from insensitivity. This interpretation not only better fits the text but has the virtue of being more charitable to Joyce. Consider the first summarizing sentence of the central passage quoted above where Joyce lays out his argument from insensitivity: “On the assumption that my favored hypothesis about the ‘moral sense’ is correct, it follows that the process by which humans form moral beliefs is an unreliable one, *for* they are disposed to do so regardless of the evidence to which they are exposed” (p. 162, emphasis mine). Joyce

Yet what, more precisely, is the standard conception of process reliability that is widely thought to be epistemically significant? Classically construed, a reliable process is just one that is truth-conducive, one that produces mostly true judgments. But after a few decades of refinement, the conception of process reliability that is fairly widely thought to be epistemically important is a modal or counterfactual one which maintains, roughly, that our judgments are reliable if they proceed from processes that produce mostly true judgments *throughout nearby worlds*, not just the actual world.³⁰ And this is widely agreed to be roughly right because distant counterfactual worlds do not seem relevant to determining the reliability of our actual belief-forming processes: if our cognitive faculties would not produce mostly true judgments in a distant evil demon world, that does not imply that we cannot rely on our cognitive faculties in the actual world.

Furthermore, the fact that our cognitive faculties produce mostly true judgments in the *actual world* does not thereby imply that they are reliable in an epistemically important sense, for our faculties may just be very lucky to get things right in the actual

appears to *infer* unreliability from insensitivity, not to identify or conflate the two together. Other passages also suggest this reading of Joyce on reliability (see pp. 160, 165).

³⁰ For a *locus classicus* of work on counterfactual process reliability, see Goldman (1986).

world in the sense that though they produce mostly true judgments in the actual world, they produce mostly false judgments in nearby worlds. So, the neighborhood of nearby worlds is the relevant range of worlds that counts for determining reliability. More precisely, then, the standard conception of process reliability runs roughly as follows:

A process type *P* is *reliable* if and only if and to the degree that its instantiations produce and/or sustain mostly true judgments throughout nearby worlds.

Correlatively:

A process type *P* is *unreliable* if and only if and to the degree that its instantiations do *not* produce and/or sustain mostly true judgments throughout nearby worlds.

Importantly, distant counterfactual worlds are not relevant for determining reliability and whether a process produces mostly true judgments in the *actual* world is not sufficient for determining its reliability.³¹

With this epistemological backdrop fresh in mind, notice that Joyce contends that a world without moral facts is relevant to the reliability and justification of our moral judgments. In that world, we would arrive at mostly false moral judgments, and thus, according to Joyce, our moral judgments are unjustified because “unreliable.” But with this move from insensitivity to unreliability Joyce faces a dilemma: he could maintain that this moral factless world is a *nearby* world or he could concede that it is a distant world yet maintain that it nevertheless is relevant for determining reliability. The latter option is dialectically ineffective, perhaps suicidal, because few (if any) epistemologists think that distant worlds are relevant for determining reliability. But the former and better option does not inspire confidence either because a world that eliminates an entire

³¹ See Becker (2008) and Henderson and Horgan (2007) for excellent motivation of the epistemic importance of this sort of actual-plus-nearby-worlds process reliability. This conception of modal reliability is widely thought to be deeply epistemically important. Many externalists think it is necessary for knowledge and/or justification. Even when it is not thought necessary, its cognitive accessibility is widely regarded by internalists and externalists as weighing significantly in favor of knowledge and/or justification. Correlatively: cognitively accessible *unreliability* is widely regarded as weighing significantly *against* knowledge and justification. In the language of epistemic defeat, having evidence of a belief-forming process’s unreliability is widely thought to provide a defeater of whatever initial justification the resulting beliefs might have had.

factual domain—the domain of moral facts— is a world that is quite distant from the actual world on standard metrics of distance or similarity. Joyce owes us an argument for why we should consider a moral factless world to be nearby. He does not offer it, and since no such argument appears ready to hand, his argument won't do. Given the dilemma, Joyce cannot secure the inference from insensitivity to unreliability.

Given that the objection from modal distance apparently disables Joyce's move from insensitivity to unreliability, what then should we make of Joyce's motivating case of the insensitivity of John's paranoid belief that was sketched above? John's belief is unreliable and unjustified because there are plausibly at least some *nearby worlds* (e.g. the actual world) where the process producing such paranoid beliefs does not produce mostly true beliefs. So Joyce's case of John and the case of our moral judgments are importantly disanalogous. Even if we grant Joycean insensitivity in both cases, Joycean insensitivity is *not* doing the debunking work in John's case: it is the unreliability of the process producing his paranoid belief, cashed out as its failure to produce mostly true beliefs in nearby worlds, which is doing the debunking work. Since the insensitivity of our moral belief-forming processes does not imply that they are unreliable—since their failure to produce mostly true beliefs in distant worlds implies nothing about their reliability—we should deny the inference from Joycean insensitivity to unreliability.

2.4 Charting Debunking Arguments from Insensitivity

Could we perhaps avoid the problem from modal distance by running an argument from the insensitivity of our *beliefs* rather than from the insensitivity of our belief-forming *processes*? Let us make the distinction:

A belief-forming *process* P is *insensitive* if and only if were the propositional contents of P 's output judgments false, P would still produce those (false) judgments.

S 's belief that p is *insensitive* if and only if were p false, S would still believe that p .

With this conceptual distinction ready to hand, we can now better chart the space of arguments from insensitivity. Specifically, we can identify three potential arguments. Joyce argues from the *process insensitivity* of all our moral judgments, and above we saw that this argument faces a formidable objection from modal distance. Another potential argument is an argument from the *joint* insensitivity of all our moral *judgments* in the sense that were there no moral facts whatsoever, we would still arrive (by the same process or a different process) at all the same moral judgments. Yet this argument falls

to the same objection from distance, since it takes an intuitively distant moral factless world to be relevant to determining the reliability of our moral judgments.

However, there is another option on the table, one that avoids the problem from distance. Perhaps Joyce could run an argument from Nozickean insensitivity, that is, the *individual* insensitivity of our particular moral *judgments*.³² That is, perhaps he could argue not that our moral belief-forming *processes* are insensitive and not that our moral judgments are *jointly* insensitive, but rather that our moral judgments are *individually* insensitive in the sense that, for each individual moral judgment, were its propositional content false, we would still arrive at the same judgment. Thus, each of our moral judgments is unjustified.³³

This argument *does* avoid the problem of distance, but the epistemic principle it relies on—the inference from the insensitivity of our individual judgments to their

³² The individual sensitivity of our beliefs of course functioned as a necessary condition in Robert Nozick's (1981) "tracking" theory of knowledge.

³³ Interestingly, Wielenberg (2010, p. 454) reconstructs Ruse's argument from insensitivity in this epistemic form, so we should seriously consider it. Wielenberg's epistemic construal of Ruse's (1986, p. 254) argument lacks textual support, but it does further the dialectic with regard to debunking arguments since it is a better argument than Ruse's actual *metaphysical* debunking argument (see Note 16 of the present chapter on Ruse). Operating with his epistemic construal, Wielenberg (pp. 454-455) critiques Ruse's move from the Explanatory Premise to the individual insensitivity of our moral beliefs, but in the next paragraph I shall challenge the move Wielenberg does not address, the move from individual insensitivity to unreliability or debunking.

unreliability or to skepticism regarding their justification—is implausible. To see this, consider the following case adapted from Linda Zagzebski:

MARY: Mary has 20/20 vision and can typically easily recognize when her husband is in the living room. This time her husband’s brother, who looks very similar to her husband, is sitting in the well-lit living room in the chair her husband usually sits in. Mary thus comes to believe that her husband is in the living room. And lo and behold, her husband is in fact in the living room, just sitting beyond Mary’s view.³⁴

Mary’s belief is insensitive in the sense that were her husband not in the living room, she would still believe he is. But her belief is nonetheless intuitively justified and the product of a reliable visual process. The Mary case impugns the move from the individual insensitivity of our moral judgments to their unreliability.

So if appeal to the insensitivity of our moral belief-forming processes does not work, and if appeal to the joint insensitivity or individual insensitivity of our moral judgments does not work either, then it appears we have exhausted our options. An argument from insensitivity to moral debunking begins to look bleak. There is, however, another argument available which takes the distinction characterized above between the joint insensitivity and individual insensitivity of our judgments and applies it to *processes* rather than judgments. The next section makes the distinction and builds

³⁴ Zagzebski (1996), p. 285. Zagzebski offered this case originally against Plantinga’s (1993) account of warrant as proper functioning. Becker (2008) also notes its usefulness as a counterexample to insensitivity principles.

upon the discussion thus far to develop a novel and much better argument from insensitivity.

2.5 A New Moral Debunking Argument from Insensitivity

This section develops a new argument from the Explanatory Premise to the Debunking Conclusion via bridge premises regarding a new species of insensitivity. First, conceptual clarification and new distinctions are in order. Both Joyce's argument from insensitivity and the argument I want to flesh out are similar in that they are arguments from *process insensitivity*:

A belief-forming process P is *process insensitive* if and only if were the propositional contents of P's output judgments false, P would still produce those judgments.

Yet our arguments are crucially distinct. On the back of the distinction between the joint insensitivity and individual sensitivity of our moral *judgments* characterized in the previous section, let us make the same sort of distinction with respect to our moral belief-forming *processes*. That is, let us make a distinction between *jointly insensitive* and *individually insensitive* belief-forming processes:

A set of belief-forming processes *S* is *jointly insensitive* if and only if were the propositional contents of *S*'s output judgments false, *S* would still produce those judgments.

A set of belief-forming processes *S* is *individually insensitive* if and only if, for each particular member belief-forming process *P*, were the contents of *P*'s output judgments false, *P* would still produce those judgments.

To appreciate the difference between the two, notice that individual insensitivity amounts to a *conjunction* of subjunctive conditionals (one conjunct for each particular belief-forming process *P*), whereas joint insensitivity amounts to a single subjunctive conditional. Joyce's argument runs on the *joint* insensitivity of our moral belief-forming processes, whereas the argument I want to develop runs on the *individual* insensitivity of our moral belief-forming processes.

Furthermore, process insensitivity, like epistemic reliability, naturally comes in *degrees*. Processes could be thought of as more or less insensitive, based on what proportion of their output judgments they would still produce were the propositional contents of those judgments false. For instance,

A set of belief-forming processes would be *completely* individually insensitive if and only if, for each particular member belief-forming process *P*, were *all* of the propositional contents of *P*'s output judgments false, *P* would still produce those (false) judgments.

My argument runs on a conception of individual process insensitivity that incorporates this degreed feature, by appealing to the *highly* individually insensitive character of our belief-forming processes. More precisely,

A set of belief-forming processes *S* is *highly individually insensitive* if and only if, for each particular member belief-forming process *P*, were **most** of the propositional contents of *P*'s output judgments false, *P* would still produce those (false) judgments.

So my insensitivity claim comes to this: for each particular moral belief-forming process *P*, were *most* of the propositional contents of its output judgments false, *P* would still produce those (false) judgments. On the basis of this claim and a few other plausible

bridge premises, we can reasonably infer the unreliability of our moral belief-forming processes. The argument runs this way:

New Debunking Argument from Insensitivity

P1: The best complete explanation of our moral judgments does not appeal to nor imply their truth. [Explanatory Premise].

P2: If (P1), then the processes³⁵ producing our moral judgments are *highly individually insensitive*.

- More precisely: if (P1), then for each particular moral belief-forming process *P*, were *most* of the propositional contents of *P*'s output judgments false, *P* would still produce those (false) judgments. On the standard analysis of subjunctive conditionals,³⁶ the claim is this: if (P1), then the *nearest worlds* where most of the propositional contents of *P*'s output judgments are false are worlds where *P* still produces those (false) judgments.

³⁵ More precisely, the *process types* whose instantiations produce our moral beliefs are highly individually insensitive. Compare: reliabilists typically identify *process types* as the relevant candidates for reliability.

³⁶ For a *locus classicus*, see Pollock (1976).

- P3: So, the processes producing our moral judgments are *highly individually insensitive*. (P1, P2)
- More precisely: So, for each particular moral belief-forming process *P*, were *most* of the propositional contents of *P*'s output judgments false, *P* would still produce those (false) judgments. On the standard analysis of subjunctive conditionals, the claim is this: So, the *nearest worlds* where most of the propositional contents of *P*'s output judgments are false are worlds where *P* still produces those (false) judgments.
- P4: There are *many* different processes producing our moral judgments.
- P5: If there are *many* different processes producing our moral judgments, then, for each particular moral belief-forming process *P*, the nearest worlds where most of the propositional contents of *P*'s output judgments are false are worlds *without a relatively small subset of the moral facts* but otherwise as similar as possible to the actual world.

- P6: Worlds without a relatively small subset of the moral facts but otherwise as similar as possible to the actual world are *nearby* worlds.
- P7: So, for each particular moral belief-forming process *P*, the nearest worlds where most of the propositional contents of *P*'s output judgments are false are *nearby* worlds. (P4-P6)
- P8: So, for each particular moral belief-forming process *P*, the nearest worlds where most of the propositional contents of *P*'s output judgments are false are both (i) worlds where *P* still produces those (false) judgments and (ii) worlds that are *nearby*. (P3, P7)
- P9. A belief-forming process is unreliable if it does not produce mostly true judgments throughout nearby worlds.
- P10. So, each of our moral belief-forming processes is unreliable. (P8, P9)

As for P1, this revision of the Explanatory Premise was motivated in section 2.2 of this chapter. As for P2, the move from the revised Explanatory Premise to the highly individually insensitive character of our moral belief-forming processes runs smoothly. For if the best complete explanation of our moral judgments does not appeal to nor

imply their truth—if the moral facts make no difference whatsoever to why we hold our moral judgments—then the processes producing our moral judgments are highly individually insensitive in the sense that for each moral-belief forming process *P*, were we to eliminate the moral facts corresponding to most of *P*'s output judgments, *P* would nevertheless still produce those (false) judgments. A natural why-question generated by this move concerns the “most” operator. Why “most” rather than “all”, given that “all” follows just as well? Alternatively put: given that the *completely insensitive* character of our moral belief-forming processes follows from the revised Explanatory Premise, why does the argument draw the weaker implication of their *highly insensitive* character? This move is made for two reasons. First, the weaker implication is all the argument needs: if in nearby worlds, *most* of process *P*'s output beliefs are false, that is sufficient for the unreliability of *P* on the standard modal understanding of process unreliability characterized earlier. Second, the weaker implication, as we shall see in the next section, nicely neutralizes two anticipated objections from the alleged necessity of moral facts.

Allow me to clarify the contrast between Joycean insensitivity and this new species of insensitivity. Joyce takes the following modal truth to have epistemic significance: were there no moral facts whatsoever, the processes producing our moral judgments would still lead us to the same judgments. In that sense, all our moral belief-forming processes are jointly insensitive. In contrast, my argument takes the following

conjunction of modal truths to have epistemic significance: for *each particular* moral belief-forming process *P*, were *most* of the moral facts corresponding to the output judgments of *P* not to obtain, then *P* would still produce the same (false) judgments. In that sense, our moral belief-forming processes are highly individually insensitive.

The *New Debunking Argument from Insensitivity*, given its new species of sensitivity, is not subject to the objection from modal distance that impugned Joyce's inference from insensitivity to unreliability. The worlds at which each process, respectively, do not lead us to mostly true judgments are *nearby worlds* where, respectively, a *relatively small subset* of the moral facts does not obtain, rather than a *distant world* where *no moral facts whatsoever* obtain. This would not be so if there were only a few process types (e.g. "evolutionary dispositions", "testimony") whose instantiations produced our moral judgments, since a world where a member of such few process types does not lead us to mostly true judgments would be a distant world in virtue of not containing a *relatively large subset* of the moral facts. But, as P4 observes, there are plausibly *many* different process types behind our moral judgments.

Two considerations motivate the claim that there many sorts of moral belief-forming processes, one empirical and one philosophical. First, in the *empirical* literature, the best explanations of our moral judgments in the evolutionary, social and psychological sciences frequently appeal to many different processes and causes (in

many different combinations), such as: the internalization of norms (an umbrella term for a variety of individuated social processes), modeling processes, operant conditioning, conformist pressures, cognitive dissonance, natural selection, affective biases, partiality bias, status quo bias, disgust reactions, and so on. So one reason to favor narrower specifications and more process types is that these better fit what the scientists are telling us about the causes of our moral judgments and are thus more warranted than armchair specifications.

Second, in the *philosophical* literature, reliabilists and others who discuss the “generality problem” —broadly, the problem of how to individuate belief-forming process types—almost always favor rather narrow specifications of processes, even though they often do not agree on how exactly to specify processes.³⁷ For example, when confronted with the question of how best to describe the type of belief-forming processes behind our ordinary visual judgments, epistemologists tend to favor narrower specifications like *processes producing visual judgments about items in the center of my visual field* over sweepingly general descriptions such as *perception* or *evolution*. One important reason given for favoring the narrower specification is that mentioned above, that they better fit what the scientists are telling us about the operative causes. Another reason is that overly general typing frequently has counterintuitive epistemic consequences. For

³⁷ For good discussion, see Becker (2008).

example, if we broadly type the processes producing *both* our visual judgments and auditory judgments as *perceptual processes*, then even if our perceptual judgments gets things right most of the time, if our auditory judgments happen to be systematically false, the counterintuitive implication is that the processes behind our visual judgments are unreliable. But the sources of our *visual* judgments, the objection goes, should not be impugned by the systematic falsity of our *auditory* judgments. So we should favor a rather narrow typing of the processes behind our ordinary visual judgments and rather narrow typing in general. All this is to say: P4 receives support from the empirical and philosophical literature.

So then, given there are plausibly *many* process types behind our moral judgments, the respective worlds where each particular process produces mostly false judgments are going to plausibly be *nearby*, since worlds without a relatively small subset of moral facts and otherwise as similar as possible to the actual world are going to plausibly be nearby. And the greater the number of process types at work, the fewer moral facts these respective worlds will be without, and so the nearer these respective worlds will be. The nearness of these worlds, as pointed out earlier, makes them relevant for determining the epistemic reliability of our moral belief-forming processes. So the fact that each particular process would produce mostly false judgments in nearby

worlds indicates, on the standard (actual-plus-nearby-worlds) conception of process reliability, that our moral belief-forming processes are unreliable.

The *New Debunking Argument from Insensitivity* thus provides us (who attend to the argument) with good evidence of the unreliability of the processes behind our moral judgments. This is a potent skeptical outcome, since justification externalists and internalists (and middle grounders) converge on the epistemic principle that having good evidence that a set of judgments is unreliably produced provides a significant defeater of any *prima facie* justification that these judgments might have had conferred upon them by moral seemings, intuitions, experience, phenomenology, or their ilk.

2.6 Three Objections

First Objection: Objection from Moral Necessity. The first objection to the *New Debunking Argument from Insensitivity* targets P6, the claim that a world without a relatively small subset of the moral facts is a nearby world. The objection presses that since the moral facts are metaphysically (or logically) *necessary* (e.g. it is wrong to torture someone just for kicks, pain is *prima facie* bad, and so on), worlds where any of them do not obtain are *impossible* worlds and you cannot get more distant than that! So, P6 is mistaken.

Response. The vast majority of the moral facts corresponding to our actual moral judgments—especially the more particular moral facts corresponding to our everyday particular moral judgments—are contingent upon empirical facts. For example, the moral fact that I wrongfully lied to someone depends on empirical facts about, say, my deceptive intention and the (actual, expected, or intended) consequences of my lying. And since empirical facts are not necessary, the vast majority of the moral facts corresponding to our actual moral judgments are not necessary. That said, there might be *some* necessary moral facts corresponding to some of our most general moral judgments or standards that do not depend on empirical facts—for example, torture just for kicks is wrong, pain is *prima facie* bad, and so on. But granting this would not impugn P6 because the counterfactual worlds in question are those where *most*, not all, of the moral facts corresponding to the output judgments of a particular moral belief forming process *P* do not obtain. Even if *P* produces *some* general moral judgments that correspond to necessary moral facts (e.g. torture just for fun is wrong), these judgments will be relatively few in number because we hold relatively few very general moral judgments compared to more particular moral judgments. So even if we grant that some necessary moral facts obtain in the counterfactual worlds in question—since they obtain in all possible worlds—that is perfectly compatible with the *nearness* of such

worlds. They are nearby worlds because they are worlds where a relatively small subset of the *non-necessary* moral facts are eliminated.

Second Objection: Objection from Counterpossibles. The second objection also springs from the alleged necessity of the moral facts and challenges the inference to P3. This objection is that, given the necessity of the moral facts, the subjunctive conditional defining the individual insensitivity of our moral belief-forming processes is a *counterpossible*—that is, a conditional with an impossible or necessarily false antecedent—and is thus, on the standard Lewis-Stalnaker semantics,³⁸ vacuously true when my argument (i.e. P3) relies on the non-vacuous truth of this conditional.³⁹

Response. This objection cannot be handled by rejecting the treatment of counterpossibles by the standard semantics.⁴⁰ For all that would do is allow for the non-

³⁸ Lewis (1973), Stalnaker (1968).

³⁹ To appreciate that the argument relies on the non-vacuous truth of the conditional defining insensitivity, consider the following. If the subjunctive conditional defining insensitivity is vacuously true, given an impossible antecedent, then the conditional defining sensitivity is also vacuously true, since it would possess an impossible antecedent as well. But the argument hinges on the claim that our moral belief-forming processes are *insensitive*—that is, *not* sensitive. For it to be true that our moral belief-forming processes are not sensitive, it must not be the case that the conditional defining individual sensitivity is vacuously true.

⁴⁰ Some argue rather cogently that the standard semantics needs to be revised somehow to allow for non-vacuously true counterpossibles. And promising revisionary proposals are available (see e.g. Vander Laan (2004)). See Nolan (1997) for an impressive argument from standard philosophical practice—namely, from the fact that we evaluate counterpossibles as

vacuous truth of the counterpossible defining insensitivity, that is, allow that even though the worlds where the antecedent holds are impossible worlds, the counterpossible is still non-vacuously true. It would *not* show the *nearness* of the worlds where the antecedent holds, but their *nearness* (to the actual world) is exactly what the argument needs to secure the inference to (actual-plus-nearby-world) unreliability. So a different response is needed.

The objection is defused by the same considerations that defused the previous objection. Since only some of the moral facts are plausibly necessary—namely, those corresponding to some of our most general moral judgments or standards—and since the subjunctive conditional defining the *highly* individually insensitive character of our moral belief-forming processes incorporates the “most” operator, the insensitivity-defining conditional in my argument is *not* in fact a counterpossible. The *New Moral Debunking Argument from Insensitivity* is not asking us to consider impossible worlds where necessary moral facts do not obtain, but rather nearby worlds where a subset of the non-necessary moral facts do not obtain. Hence the *Objection from Counterpossibles* does not get off the ground.

true all the time in philosophy—for thinking the standard semantics should be revised to allow for non-vacuously true counterpossibles.

Third Objection: Objection from Moral Supervenience. This objection maintains that the truth of moral supervenience challenges P6 and P1. Consider P6—the premise that worlds without a relatively small subset of the moral facts but otherwise as similar as possible to the actual world are *nearby* worlds. The objection presses that P6 is only plausible on the assumption that the moral facts are metaphysically independent of their underlying *natural facts*, such that the removal of various subsets of the moral facts would *not* entail the removal of their underlying natural facts as well. That is, P6 is only plausible on the assumption that moral supervenience does not obtain. For if moral supervenience does obtain—and it is widely held to obtain in some form or other—this casts doubt on the claim in P6 that worlds without a relatively small subset of the moral facts are *nearby*, since they are not just worlds where a subset of moral facts do not obtain but rather also worlds where the natural facts that underlie these moral facts do not obtain.

Moral supervenience also appears to undermine the plausibility of P1, the revised Explanatory Premise that the best explanation of our moral judgments does not appeal to *nor imply* their truth. This is so because it is plausible to suppose that the natural facts underlying moral facts frequently play a role in the best explanation of our moral beliefs. For example, the natural facts that constitute torture play a role in explaining why we think torture is wrong. If so, then if moral supervenience holds such

that certain natural facts necessitate or imply certain moral facts, the best explanation of our moral beliefs would frequently *imply* their corresponding moral facts. So moral supervenience also appears to undermine P1.

Response. Consider the sort of moral supervenience at issue:

Supervenience: It is impossible⁴¹ for moral facts to differ without there also being a difference in natural facts, but not vice versa.

Crucially, we must distinguish a stronger and weaker version of the claim.

Strong Supervenience: It is impossible for moral facts to differ *between any two possible worlds* without there also being a difference in natural facts, but not vice versa.

⁴¹ Some proponents of moral supervenience accept it as a semantic or analytic truth regarding the meaning of moral terms (or as a requirement on the competent use of moral terms), whereas others accept it as a non-analytic truth about relations of metaphysical necessity holding between two types of properties. The difference does not affect the discussion that follows.

Weak Supervenience: It is impossible for moral facts to differ *within one possible world* without there also being a difference in (interesting⁴²) natural facts, but not vice versa.

Let us then ask: does the objection from moral supervenience rely on Strong Supervenience or Weak Supervenience? It appears to rely on Strong Supervenience, since the objection has no force if we instead accept only Weak Supervenience. To appreciate this, consider the objection's main point: it is that in a possible world, the fact that a subset of the moral facts would not obtain *entails* that a lot of natural facts in that world would also not obtain. But apparently this would only be so if certain natural facts *necessarily* underlie a certain moral fact, that is, underlie it in *all* possible worlds where the natural facts obtain. For example, on Strong Supervenience, if a naturalistically specifiable instance of you lying to me is morally wrong, then the same

⁴² Two distinct actions (or motives, institutions, etc) within a possible world obviously will not have the *same* natural properties—after all, they cannot be identical in terms of place if they are distinct. The Weak Supervenience claim is not the trivial truth that within a world if we keep *all* the natural facts identical, the moral facts must be identical in that world. Of course that holds. The interesting claim, rather, is that setting aside *uninteresting* differences in natural facts—e.g. differences in the place of an action or the numeric identity of the agent—then if we fix the *interesting* natural facts in a world, then there must be no difference in the moral facts in that world. For instance, if suicide is always wrong in a world, then if John and Bill are both identical in all interesting naturalistic respects and commit suicide in the same way (i.e. two suicides that differ only in being committed by different agents in different places), their suicides must both be wrong in that world.

naturalistically specifiable act is wrong in every possible world where it obtains.

However, if certain natural facts necessarily underlie certain moral facts only *within the same possible world*—as Weak Supervenience claims—then the fact that a subset of the moral facts do not obtain would *not* entail that the underlying natural facts in that world do not obtain. For example, on Weak Supervenience, if a naturalistically specifiable instance of you lying to me is morally wrong, it is an open question whether the same naturalistically specifiable instance of lying would be wrong in all possible worlds where it obtains.

The discussion of moral supervenience draws out an important scope limitation of the *New Moral Debunking Argument from Insensitivity*. That is, the new argument appears not to work against anti-skeptical metaethical views that happen to accept Strong Supervenience.⁴³ Does this limitation in scope zap the new argument's

⁴³ The new argument from insensitivity also may not work against a possible moderate or intermediate form of moral supervenience where certain natural facts underlie certain moral facts throughout the range of nearby worlds where the natural facts obtain, rather than throughout all possible worlds where the natural facts obtain. More precisely, moderate moral supervenience would claim: it is impossible for moral facts to differ *between any two worlds within the range of nearby worlds* without there also being a difference in (interesting) natural facts, but not vice versa. To my knowledge, this possible moderate form of moral supervenience is neglected by the metaethical literature and its motivations are not apparent: why would certain natural facts underlie certain moral facts only through the range of nearby worlds rather than in all possible worlds where the natural facts obtain? First look, the cut off point appears rather arbitrary. That said, if a version of the moderate supervenience suggested allows moral anti-skeptics to avoid the *New Debunking Argument from Insensitivity*, that could be seen as one of its motivations. So the success of the *New*

philosophical interest? Does it imply that the new argument leaves the metaethical landscape as it was before? The following reasons suggest not.

First, observe that though Weak Supervenience is widely held, it is not at all clear that Strong Supervenience is widely accepted or should be accepted in the first place⁴⁴ There is a notable dearth of arguments for it. The most prominent sort of argument for Strong Supervenience moves from the alleged *inconceivability* of moral differences without natural differences to the *impossibility* of such differences.⁴⁵ But such arguments

Debunking Argument from Insensitivity could affect the metaethical landscape by pressing moral anti-skeptics toward developing and accepting moderate forms of moral supervenience rather than just Weak Supervenience. For the purposes of this paper, however, since moderate forms of moral supervenience are as of yet undeveloped and unmotivated in the metaethical literature, nothing more will be said about them here.

⁴⁴ Some ethical naturalists accept Strong Supervenience—notably some non-reductive ethical naturalists (e.g. Cornell Realists such as Brink (1989, p. 175)) and analytic naturalists (e.g. Canberra Moral Realists such as Jackson (1998, pp. 130-31))—but other ethical naturalists do not (e.g. many “constructivists” who hold that the moral facts are constituted by our (actual or ideal) attitudes or responses do not accept Strong Supervenience). Though some ethical non-naturalists (e.g. Shafer-Landau (2003)) accept Strong Supervenience, they need not do so and may be better off not doing so (see Hills (2009) for this case). To my knowledge, nearly all these major metaethical camps accept Weak Supervenience. Moreover, many noncognitivists (e.g. Blackburn (1971, 1986)) accept Weak Supervenience but reject Strong Supervenience. Interestingly, however, Hills (2009) formidably argues that we should reject Strong *and* Weak Supervenience and settle for something yet weaker, what she calls *Constant Conjunction*: “in the actual world, there are no differences in moral properties without differences in natural properties” (p. 167).

⁴⁵ See e.g. Shafer-Landau (2003, pp. 78) and Ridge (2008).

have a difficult time vindicating the alleged “inconceivability” and the inference from inconceivability to impossibility.⁴⁶

Second, others have pointed out that Strong Supervenience demands an explanation if we are to take it seriously (it is not plausibly a brute fact), since it amounts to the rather bold postulation that a metaphysical relation of necessity holds between two types of properties. Given that plausible explanations of Strong Supervenience are not yet at hand,⁴⁷ the commitment to Strong Supervenience, far from being a metaethical nonnegotiable, appears to be more of a theoretical cost.⁴⁸

⁴⁶ See Hills (2009, sect. 5) for a formidable critique of arguments for Strong Supervenience that run from inconceivability to impossibility.

⁴⁷ See Ridge’s (2007) critique of Shafer-Landau’s (2003) non-naturalist explanation of Strong Supervenience. Reductive ethical naturalists, non-reductive ethical naturalists (e.g. Cornell Realists like Brink (1989)) and analytic naturalists (e.g. Canberra Moral Realists like Jackson (1998)) appear to have an easy explanation, for on their views the moral facts are identical to certain natural facts and since all facts supervene upon themselves, it is easy to see why the moral facts supervene upon natural ones. But Ridge (2008) cogently argues that if we construe Strong Supervenience more perspicuously and more uncontroversially in terms of the supervenience of the moral on the *non-moral* (rather than the “natural” *qua* the facts and properties that the sciences investigate), then these ethical naturalists also possess an explanatory burden that is not easily met.

⁴⁸ For argument along these lines, see Mackie, (1977), Blackburn (1971, 1988), and Hills (2009).

To sum up: the *Objection from Moral Supervenience* usefully shows that *New Debunking Argument from Insensitivity* does not apply to anti-skeptical metaethical views that accept Strong Supervenience. But since many metaethical accounts do not accept Strong Supervenience and since it is at least unclear whether they should do so, the success of the *New Debunking Argument from Insensitivity* will affect the metaethical landscape. How it affects one's overall metaethical view, of course, depends on what else one holds. For example, it will press metaethical views that do not accept Strong Supervenience toward moral justification-skepticism. Or it will press metaethical views to accept and defend Strong Supervenience in order to avoid the skeptical conclusion.⁴⁹ Regardless, the success of the *New Debunking Argument from Insensitivity* will not leave metaethics where it was before.

2.7 Concluding Remarks

The move from the Explanatory Premise to the Debunking Conclusion does not inspire confidence: it is opaque how to get from here to there. But the *New Debunking Argument from Insensitivity* developed in this chapter shows how we can add a few plausible bridge premises to the Explanatory Premise in order to reasonably get to the

⁴⁹ Or it will press metaethical views to develop and support a neglected moderate form of supervenience in order to avoid the skeptical conclusion. See Note 43 of the present chapter.

Debunking Conclusion. The payoff is that the plausibility of moral anti-skepticism hinges on the merit of the Explanatory Premise and thus on the merit of moral explanations. This outcome gives a renewed sense of urgency to the debate over moral explanations. Moreover, it lends support to metaethical accounts that can vindicate moral explanations over metaethical accounts (e.g. robust ethical non-naturalisms) that apparently have a much harder time doing so. If all otherwise plausible metaethical accounts cannot vindicate moral explanations, then we may be pressed toward moral justification-skepticism full stop.

The dialectical situation is quite philosophically pregnant because many hold that the Explanatory Premise applies to many or all of our moral beliefs, and, moreover, many hold that it applies to many or all of our philosophical and religious beliefs. If they are onto something, then the argument developed in this chapter pushes us toward a debunking conclusion regarding such beliefs. If the *New Argument from Insensitivity* (or some modification of it) works, the skeptical implications could be rather staggering.

Chapter 3: Street's Darwinian Dilemma and Enoch's Epistemological Challenge

3.1 Introduction

This chapter explores prominent ways of fleshing out a provocative and almost completely neglected sort of debunking challenge, namely an explanatory challenge to moral realism that runs from considerations concerning the origins of our moral judgments. The challenge to the moral realist is essentially this: explain moral reliability or face moral skepticism. The challenge is motivated by the difficulty of explaining why we should trust or rely upon the evolutionary, socio-cultural, and psychological processes behind our moral judgments to lead us to objective (response-independent) moral truth. The central premise is that realists have no plausible explanation of moral reliability and the conclusion is moral skepticism.

The explanatory challenge can morph into different forms with different scopes. Giving recent and rare attention to the challenge, Sharon Street (2006) and David Enoch (2010) develop sophisticated versions of it that possess different scopes. Enoch targets a robust or platonic form of moral realism with his "Epistemological Challenge", while Street with her "Darwinian Dilemma" targets robust and non-robust versions of moral realism that accept the response-independence of the moral facts.

Enoch's (2010) Epistemological Challenge, which he self-consciously develops on the shoulders of Street's (2006) Darwinian Dilemma, has yet to receive critical treatment. Street's formidable Darwinian Dilemma has impressed prominent figures and inspired recent defenses of moral realism.¹ In a recent response, Copp (2008) attempts to evade Street's challenge by arguing that there is in fact a plausible explanation of moral reliability, namely one relying on his own "society-centered" version of ethical naturalism. Similarly, realists could respond to Enoch's Epistemological Challenge—as Enoch himself ultimately does —by developing plausible explanations of moral reliability. But I argue later in the next chapter that realist explanations of moral reliability—including Copp's (2008) explanation and Enoch's (2010) explanation—are implausible. This raises the question: if realist explanations are implausible, should we then accept the cogency of Street's and Enoch's explanatory challenges? No, or so I argue in the present chapter. That is, I argue that *even if we grant* that realists possess no plausible explanation of moral reliability, prominent explanatory challenges that run from this premise are unsuccessful. Why? Because these challenges share a central defect, namely the poor quality of their central inference to moral unreliability from the

¹ Allan Gibbard (2011) suggests that Street's Darwinian Dilemma succeeds against moral realism but argues that it does not succeed against his own quasi-realist metaethics. Similarly, Copp (2008) motivates Street's Darwinian Dilemma but contends that it does not impugn his own "society-centered" ethical naturalism. Fitzpatrick (2008), Kramer (2009), and other moral realists have also defended moral realism against Street's Darwinian Dilemma.

premise that realists have no plausible explanation of moral reliability. The fact that they do not secure this inference not only demonstrates the inadequacy of prominent explanatory challenges but usefully and positively guides us to a new and better explanatory challenge which secures (or at least better secures) the inference. The subsequent two chapters—Chapters 4 and 5—then develop and back up this new explanatory challenge.

3.2 Enoch's Epistemological Challenge

Start with Enoch's (2010) Epistemological Challenge. He targets *robust moral realism*, which he characterizes in terms of two claims:

Metaphysical Robustness: the moral facts are causally inert, response-independent, non-spatiotemporal facts.

Correlation: most of our actual moral beliefs are (non-accidentally) true.

Alternatively put: there is a counterfactually stable or reliable correlation between our moral beliefs and the moral facts.

Enoch's official target is "robust meta-normative realism" or robust realism about value or normativity in general—regarding practical reasons, moral reasons,

epistemic reasons, and so on—rather than robust realism about *moral* value or normativity in particular. But for reasons of economy and continuity, let us focus on robust *moral* realism.

By construing his target in terms of the robustness of the moral facts, Enoch is not targeting ethical naturalisms or non-naturalisms that accept the causal efficacy of the moral facts, nor is the challenge targeting response-dependence views of the moral facts. Rather, his target is a type of non-naturalist moral realism that denies the causal efficacy of the moral facts, a kind of moral platonism.²

Enoch targets *robust* moral realism because it is especially vulnerable to the challenge he develops and also, importantly, because it is ultimately the metaethical view he himself wants to defend in his target article (2010) and in his recent (2011) book. His target article is structured as follows. He first develops what he calls an “Epistemological Challenge” to robust moral realism that runs from the premise that robust moral realists have no plausible explanation of moral reliability. Enoch then tries to defuse the challenge developed by arguing that there is in fact a plausible explanation

² It is important to make a distinction between robust moral realism and non-naturalist moral realism, for though non-naturalist moral realists tend not to accept the causal efficacy of moral facts, they may consistently accept their causal efficacy and may even be better off doing so, as Majors (2007) suggests. This is to point out that if Enoch’s challenge succeeds against robust moral realism that does not imply it succeeds against *all* versions of non-naturalist moral realism.

of moral reliability available, namely his own “evolutionary pre-established harmony” explanation.³ For present purposes, set Enoch’s own explanation to one side (we assess it later in the next chapter). The present task is to assess the cogency of Enoch’s challenge. After getting clear on his challenge, I develop an objection that demonstrates its inadequacy even if we grant that robust moral realists in fact have no plausible explanation of moral reliability. The objection shows the success of Enoch’s challenge does not in fact depend, as Enoch suggests it depends, on whether he or any other robust moral realist has a plausible explanation of moral reliability.⁴

With target identified—robust moral realism—move with me to the reconstruction and motivation of Enoch’s challenge. Inspired by Hartry Field’s Benacerrafian epistemological challenge to mathematical Platonism,⁵ Enoch raises the following question: the robust moral realist claims there is a reliable correlation between

³ Enoch (2010), p. 430. For the explanation, see section 5.3 of Enoch’s paper (pp. 430-35).

⁴ In Chapter 4, I argue that moral realists (robust varieties included) have no plausible explanation of moral reliability. This does not vindicate Enoch’s challenge, however, since his challenge falters on other grounds, as the present chapter will show.

⁵ See Benacerraf (1973) and Field (1989, pp. 25-30, 230-39). Field argues that the fact that mathematical facts, on Platonist views, are non-spatiotemporal and causally inert generates *prima facie* reason to think that the mathematical Platonist cannot plausibly explain the reliability of our mathematical beliefs, that is, cannot plausibly explain how the processes or mechanisms producing our mathematical beliefs tend to produce mostly true ones. Without a plausible explanation of mathematical reliability available, Field suggests that we have *prima facie* reason for thinking that our mathematical beliefs are unreliable on Platonist models of the mathematical facts.

our moral beliefs and the robust moral facts, *but what plausibly explains this striking correlation?* That is, even if we grant the existence of moral facts and the robust moral realist's understanding of their nature, how did we wind up with so many true moral beliefs if the moral facts that correspond to these beliefs are "outside" of space and time,⁶ causally inert,⁷ and independent of all human responses and attitudes? Surely we must have an explanation if we are to reasonably accept such a marvelous coincidence; otherwise, it appears too damn good to be true.

Enoch frames the challenge most perspicuously in the following two central passages:

...[H]ere is the version of the epistemological challenge I suggest that we focus on: Very often, when we accept a normative judgment *j*, it is indeed true that *j*; and very often when we do not accept a normative judgment *j* (or at least when we reject it), it is indeed false that *j*. So there is a correlation between (what the realist takes to be) normative truths and our normative judgments. What explains this correlation? On a (robustly) realist view of normativity, it can't be that our normative judgments are causally or constitutively responsible for the normative truths, because robust realism is inconsistent with a response-dependence view of normativity, and so the normative truths are supposed to be independent of our normative judgments. And given that (at least basic) normative truths are causally inert, they are not causally responsible for our

⁶ The notion of moral facts being "outside" anything is of course a spatial notion. The idea that the moral facts are "outside" of spacetime is just the idea that they exist but not in spacetime (analogously: God, traditionally conceived, exists but not in spacetime).

⁷ A fact is causally inert or inefficacious if it stands in no causal relations to us or anything else.

normative beliefs. Nor does there seem to be some third-factor explanation available to the realist. And so the [normative] realist is committed to an unexplained striking correlation, and this may just be too much to believe. (Enoch (2010), pp. 421-22)

Given the apparent unavailability of any such explanation to the robust [normative] realist, then, the challenge concludes that such realism is highly implausible....[O]ur challenge, then: If a brute correlation is too much to believe, and if no explanation is available to the realist, then the realist must conclude that there is after all no correlation between his normative beliefs and the normative truths. And knowing that, any (internalistically understood) justification he may have had for his normative beliefs is defeated (or undermined). (Enoch (2010), p. 424)

The demand is not for a *possible* explanation—that would be too easy.⁸ Nor does Enoch's challenge demand *the correct and complete* explanation—that would be too hard.⁹ Rather, the demand is for a *plausible* explanation of the correlation—roughly put, one that is likely to be true. Furthermore, the demand is not necessarily for a *causal* explanation. It is a demand is for a plausible explanation of whatever sort available, causal or non-causal.

⁸ For example, robust moral realists could say it is possible that we just reliably “see” or intuit the moral facts in a quasi-perceptual way even though we have no contact with them. Of course, if we have no contact with the moral facts, it is very hard to see how we could reliably see or intuit them. So this possible explanation does not appear to be plausible.

⁹ The demand is too strong because we apparently do not have *the correct and complete* explanation of the reliability of our *perceptual* beliefs, though we certainly have *plausible* explanations in broad outline in terms of our evolved psychological apparatus interacting with the external environment.

We have in hand a plausible enough explanation of the reliable correlation between the perceptual facts and our perceptual beliefs, at least in broad outline. If we grant the existence of perceptual facts and a standard realist understanding of their nature—that is, if we grant the objective existence of photons, optic nerves, external objects, the natural selection of our visual equipment—then we (or the relevant scientists) can sketch a nice causal story that makes good sense of how we have wound up with so many true perceptual beliefs. But apparently no such perceptual (or quasi-perceptual) story is available in the moral case because the robust moral facts, unlike perceptual facts, are not causally related to our minds or anything else for that matter.

Enoch's explanatory demand should bother robust moral realists because the standard ways of explaining correlations between two factors apparently do not work in the case of the correlation between our moral beliefs and the moral facts if the moral facts are construed so robustly. Given the causally inert character of the moral facts, there *cannot even in principle* be a causal or mechanistic connection—as there plausibly is in the case of our perceptual and scientific beliefs—between the moral facts and our moral beliefs that might explain why they are correlated. Moreover, given the response-independent character of the moral facts, there *cannot even in principle* be a constitutive connection—as perhaps there is in matters of aesthetics or etiquette—between the moral facts and our moral beliefs that might explain why they are correlated. Nor does there

appear to be any such explanatory connection between these two factors and some common cause or third factor that, in virtue of being “responsible” (causally or constitutively) for these two factors, might explain why they are so correlated. Importantly, such a “third-factor explanation” is not unavailable in principle, but it does appear to be unavailable *at present*. For instance, though it is *possible* that (say) the Leibnizian God—as the “third factor”—coordinates or pre-establishes the harmony between our moral beliefs and the moral facts, this is not a convincingly plausible explanation of their correlation because, for one thing, it posits the existence of the Leibnizian God.¹⁰

So the robust moral realist finds herself in an awkward situation: she is committed to a rather striking correlation for which she apparently has no plausible causal or non-causal explanation, given that she cannot even in principle invoke causal or constitutivist explanations and given that there are apparently no plausible third factor explanations available.

¹⁰ The Leibnizian explanation suggested here piggybacks off Leibniz’s doctrine of the pre-established harmony between mind and body. Leibniz’s doctrine holds that God is causally responsible for the fact that we have a mind and body and coordinates the mind and body to act *as if* they causally interacted (if you stomp on my toe, I typically feel pain) even though they really do not. Similarly, the Leibnizian explanation of the correlation between our moral beliefs and the moral facts runs as follows: God is causally responsible for our moral beliefs and is constitutively responsible for the moral facts and, moreover, coordinates our moral beliefs and the moral facts such that we form our moral beliefs *as if* we are causally responding to the moral facts even though we really are not.

To get clear on the challenge, an important interpretative question must be resolved: what does Enoch mean by “correlation” in saying there is no plausible explanation of the correlation between our moral beliefs and the robust moral facts? It is clear the correlation the robust moral realist posits is not *perfect*—we make mistakes. Nonetheless the idea is that we (at least some of us) form a sufficiently high percentage of true moral beliefs. As Enoch puts it, the targeted robust moral realism holds that our moral beliefs are “reliable.” But is this conception of reliability restricted to the actual world or is it a counterfactual (or “nonaccidental”) conception that has implications for nearby worlds as well? In his paper, Enoch characterizes reliability in both ways: that is, as actual world reliability (most of our moral judgments are true)¹¹ and as nonaccidental reliability (it is not an accident that most of our moral judgments are true).¹² But he

¹¹ See e.g. Enoch (2010), p.421:

...[H]ere is the version of the epistemological challenge I suggest that we focus on: Very often, when we accept a normative judgment *j*, it is indeed true that *j*; and very often when we do not accept a normative judgment *j* (or at least when we reject it), it is indeed false that *j*. So there is a correlation between (what the realist takes to be) normative truths and our normative judgments. What explains this correlation?

¹² We find evidence of this in an important footnote. In this note Enoch identifies the conception of reliability he operates with in his formulation of the explanatory challenge:

You may think that more [than the truth of a “sufficiently large portion” of a class of beliefs] is needed for reliability, and in particular that reliability contains some modal feature, so that a set of beliefs is reliable only if, roughly speaking, *it is not an accident* that sufficiently many of them are true. At least one natural way of understanding

makes it crystal clear at one point that it is nonaccidental reliability that figures in his challenge.¹³ In other words: it is the *counterfactually stable* correlation between our moral beliefs and the robust moral facts that matters for Enoch. And rightly so: after all, if the robust moral realist were only committed to an actual world correlation, then though our moral beliefs be in fact mostly true, this may be a complete accident of the sort that is incompatible with the justificatory or knowledge status of these beliefs.

With a grip on the content and motivation of Enoch's claim that there is no explanation of the correlation between our moral beliefs and the robust moral facts, we can spell out the rest of his challenge. Enoch lays it out concisely in this passage:

[O]ur challenge, then: If a brute correlation is too much to believe, and if no explanation is available to the realist, then the realist must conclude that there is after all no correlation between his normative beliefs and the normative truths. And knowing that, any (internalistically understood) justification he may have had for his normative beliefs is defeated (or undermined) (Enoch (2010), p.424)

such a requirement leads to what I take to be the proper understanding of the epistemological challenge, in Sect. 3. (Enoch 2010, p.148, n. 12, emphasis mine)

¹³ See quote in previous Note.

According to Enoch, the counterfactually stable correlation between our moral beliefs and the moral facts is probably¹⁴ not a metaphysically brute, unexplainable fact like (say) the existence of the universe. Given that the correlation, if it exists, must have some explanation or other, the fact that we have no plausible explanation of it gives us good reason to think that *there is in fact no such correlation*. The basic idea behind the inference is that given that the go-to explanatory games in town—causal and constitutivist explanations—are not even in principle available to the robust moral realist and given that there apparently are no plausible third-factor explanations available even in broad outline, the robust moral realist should seriously reconsider whether the *explanandum* she is committed to actually obtains. Enoch’s point is not just the purely negative one that the robust moral realist has no plausible explanation, but that the above considerations regarding how the standard explanations of correlations are *in principle* ruled out and how nonstandard “third-factor” explanations are apparently unavailable even in broad outline give us *positive reason* for thinking there is no explanation to be had. If we have good reason for thinking there is no explanation of the correlation to be had, then given the correlation is probably not a metaphysically brute, unexplainable fact, we have good reason for thinking there is no correlation.

¹⁴ Enoch’s talk of the correlation being “highly unlikely” to be brute appears merely to be his way of saying there is good reason to think the correlation is not brute. Nothing more (e.g. like the appeal to frequency of brute correlations) appears to be packed into the claim.

The suspicion could arise that Enoch's challenge relies on contentious externalist premises—that is, that it relies on premises that assume, roughly, that justification or knowledge does not require access to or awareness of the epistemic grounds. But Enoch points out that his challenge is not purely externalist, since it can easily be “internalized” for those of us who attend to it. That is, we who run Enoch's argument in our heads can come to awareness and justified belief (internalistically understood) that there is no plausible explanation of moral reliability and therefore that our moral beliefs are unreliable (on robust moral realism). And if the robust moral realist who is aware of Enoch's argument is pressed to conclude she is justified (internalistically) in believing that her moral beliefs are unreliably formed, then, regardless of whether she is an externalist or internalist, she thereby has a defeater of the justification of all her moral beliefs that defeats whatever initial justification they enjoyed, that is, the robust moral realist is thereby pressed to conclude that she is not justified in holding her moral beliefs. Enoch is not claiming that *de facto* non-accidental reliability is *necessary* for justification—for that would be a contentious externalist claim and Enoch self-consciously tries to steer clear of contentious epistemic claims—but rather that (internalistically) justified belief in non-accidental unreliability defeats justification. Enoch rightly points out that the vast majority of epistemic externalists, internalists, and

middle grounders alike would converge on this epistemic principle.¹⁵ Given the fairly neutral status of the underlying epistemic principles, Enoch's challenge appears forceful.

One final step: if the assumption of robust moral realism propels us to moral justification-skepticism, then by Enoch's lights this is a big mark, if not a completely fatal mark, against that assumption. If pressed to justification-skepticism the robust moral realist

can then maintain his commitment to realism only at the price of a rather thoroughgoing skepticism about the normative. And while this is a possible position to have, stakes have certainly been raised: If the only way to be a realist is to deny epistemic justification for any normative belief (at least of the epistemologically informed, those who can run the [challenge] in their heads), then antirealism gains significant ground. (Enoch (2010), p.424)

Enoch's challenge concludes that a realist moral metaphysics is "highly implausible," since it implies moral justification-skepticism.¹⁶

¹⁵ Of course, some extreme externalists may refuse to get on board. They might insist that since (internalist) justification can come apart from *de facto* reliability, then even though we may be (internalistically) justified in believing in moral unreliability, it still could be that *de facto* moral reliability holds and thus that our moral beliefs are justified since *de facto* reliability is sufficient for justification. However, most externalists, impressed by clairvoyant cases (or counterparts), supplement bare *de facto* reliability with a "no defeater" condition in their accounts of justification and count good (internally construed) evidence of unreliability as a defeater of justification. For instance, if I (internalistically) know (or justifiably believe) that my visual beliefs are unreliable in poorly lit conditions, if I know that I have formed a particular set of visual beliefs under poorly lit conditions, and if I thereby know that these visual beliefs are unreliable, then I have excellent reason to think I am not justified in holding them.

With the challenge laid out, a couple clarifications should be made. Importantly, Enoch's challenge is not an argument for moral justification-skepticism on *all* metaethical accounts of the moral facts because some versions of ethical naturalism—namely, realist versions that buy the response-independence of the moral facts—can *at least in principle* meet the challenge by showing how the moral facts *qua* natural facts have been causally involved in the production of our moral beliefs in a way that would allow us to plausibly explain why most of our moral beliefs are reliable. The expected analogy would be with perceptual facts: perceptual facts play a role in the best explanation of our perceptual beliefs and thus play a role in plausible explanations of perceptual reliability.

Moreover, response-dependent theorists, sensibility theorists, metaethical “constructivists” and their ilk who posit a *constitutive* connection between the moral facts and our moral judgments (or dispositions or other attitudes) may also *at least in principle* explain why they are correlated. For instance, a simple response-dependence view that holds the moral facts are constituted by our actual moral judgments has an easy explanation of moral reliability: our moral judgments are reliably true because our

¹⁶ Enoch (2010), p. 423.

holding them *makes them* true and we could not have easily had very different moral judgments. A more sophisticated response-dependence view could hold that the moral facts are constituted by our slightly idealized (e.g. fully informed) moral judgments. The explanation of moral reliability on this metaethics would then run: our moral judgments are reliably true because we reliably satisfy the relevant idealized conditions in holding our moral judgments.

But why does Enoch's challenge target *robust* moral realism as opposed to moral realism more generally? For two reasons. First, as suggested earlier, as a dialectical matter Enoch's challenge targets robust moral realism because it is the view that he ultimately wants to defend. Second, though he recognizes that his challenge *might* impugn some realist versions of ethical naturalism¹⁷ and even some *ideal* response-dependence views¹⁸ on which there is not a plausible explanation of the correlation between our moral beliefs and the moral facts, robust moral realism is *specially* targeted because, given that its claims of Metaphysical Robustness and Correlation preclude causal and constitutivist explanations of moral reliability, there is positive reason for

¹⁷ Enoch (2010), p.422.

¹⁸ Enoch (2010), p. 423, n. 20.

thinking no version of robust moral realism can provide the demanded explanation.¹⁹ In other words, robust moral realism is vulnerable to Enoch's challenge in a way that other metaethical views are not, much like mathematical Platonism is thought to be especially vulnerable to the Benacerrafian challenge. Non-robust versions of moral realism and even idealized response-dependence views are not out of the woods yet either, but at least they have access *in principle* to causal and constitutivist explanations of moral reliability. So Enoch's challenge does not apply with equal force to all metaethical views.

For clarity and ease of reference, Enoch's challenge can be laid out step-wise:

Enoch's Epistemological Challenge

E1: The alleged correlation posited by the robust moral realist between our moral beliefs and the robust moral facts is highly unlikely²⁰ to be metaphysically brute, unexplainable.

¹⁹ Enoch (2010), p. 422.

²⁰ Again, Enoch's talk of the correlation being "highly unlikely" to be brute appears merely to be his way of saying there is good reason to think the correlation is not brute. Nothing else (e.g. an appeal to frequency likelihood like "very few correlations are brute") is packed into the claim.

E2: There are three available types of explanations for the alleged correlation: causal, constitutivist, and third-factor explanations.

E3: The robust moral realist, in principle, has no causal explanation of the alleged correlation.

E4: The robust moral realist, in principle, has no constitutivist explanation of the alleged correlation.

E5: The robust moral realist also has no plausible “third-factor” explanation of the alleged correlation.

E6: So, the robust moral realist has no plausible explanation of the alleged correlation. (E2-E5)

E7: So, we²¹ are justified in believing there is no correlation (i.e. our moral beliefs are unreliable on robust moral realism). (E1, E6).

E8: If we are justified in believing there is no correlation (i.e. that our moral beliefs are unreliable), then we are not justified in holding our moral beliefs (on robust moral realism).

E9: So, we are not justified in holding our moral beliefs (on robust moral realism). (E6-E8)

E10: So, robust moral realism is highly implausible. (E9)

²¹ “We” refers to those of us who attend to Enoch’s argument in our heads.

This reconstruction accurately captures and fleshes out Enoch's most explicit statement of the argument (as the inserted brackets show):

Given the apparent unavailability of any such explanation to the robust [normative] realist, then, the challenge concludes that such realism is highly implausible [E10]....Getting back to our challenge, then: If a brute correlation is too much to believe [E1], and if no explanation is available to the realist [E2-E6], then the realist must conclude that there is after all no correlation between his normative beliefs and the normative truths [E7]. And knowing that, any (internalistically understood) justification he may have had for his normative beliefs is defeated (or undermined) [E8-E9]. (Enoch (2010), p. 423-424)

It is worth repeating that, in the final analysis, Enoch thinks this challenge does not succeed because he thinks there is in fact a plausible third factor explanation of the correlation between our moral beliefs and the robust moral facts, namely his “(godless) pre-established-harmony” evolutionary explanation.²² He develops this explanation later in the same article in which he lays out the explanatory challenge reconstructed above. Again, for the purposes of the present chapter, we set Enoch's evolutionary explanation one side (we assess it later in the next chapter). The present task now that we understand Enoch's challenge is to assess its cogency. I argue in the next section that it is not cogent, owing to the defectiveness of its central skeptical inference from the premise that we have no explanation of reliability. This section

²² Enoch (2010), p. 430.

shows that Enoch's challenge can be defused even if there is in fact no plausible explanation of the correlation between our moral beliefs and the robust moral facts and thus that the cogency of his challenge does not hinge on whether such an explanation can be provided by him or anyone else.²³

3.3 Objections to Enoch's Inference from No Good Explanation

My objections to Enoch's Epistemological Challenge target his central inference from E1 and E6 to E7, the move from the claims that the correlation is not metaphysically brute (E1) and that we have no plausible explanation of the correlation (E6) to the conclusion that we are justified in believing there to be no correlation (E7).

The inference does not work, or at least Enoch does not show how it works or could work. The notable chasm in the inference is brought out by consideration of a case that undermines the general principle apparently underlying the inference, that is, the general principle that if the correlation between our beliefs and the corresponding fact is not brute and we have no plausible explanation of the correlation—in the sense specified by Enoch that the standard forms of explanation (namely, causal and constitutivist explanations) are ruled out in principle and we have no plausible “third-

²³ See Note 4 of the present chapter.

factor” explanations available—then we are justified in thinking there is no such correlation. The case:

MATHEMATICAL RELIABILITY: There is a possible world where we are justified in believing (e.g. via considerations given by universality, self-evidence, or indispensability for scientific theorizing) that at least some of our mathematical beliefs are reliable (on platonist understandings of the mathematical facts) even though we have no plausible causal, constitutivist, or third-factor explanation of their reliability.

For many mathematical platonists, not only is such a world possible, such a world is the actual world! Nonetheless, if skeptical about its actuality, notice that its force as a counterexample does not hinge on its actuality. That this world is possible shows that E7 does not follow directly from E1 and E6. It shows that just as we can justifiably believe *that* something is the case without being able to explain *why* it is the case, we can justifiably believe that our moral beliefs are reliable without being able to offer a plausible explanation of that reliability, even if we grant that causal and constitutivist explanations of that reliability are in principle unavailable and third-factor explanations are presently unavailable. For example, accessible considerations given by the universality of our mathematical beliefs, their self-evidence, and their indispensability to scientific theorizing could possibly ground justified belief in mathematical reliability even if we do not yet have a plausible explanation of that

reliability in hand. The case suggests that we cannot reasonably infer reliability skepticism in the way Enoch suggests.

However, perhaps we could cast Enoch's inference as an inference to the best explanation (IBE). He does not cast it this way or argue for it this way, but we could do so and see where it gets us. The structure of the new IBE argument would then run as follows. First step: if we have no plausible explanation of the correlation (E6) and the correlation is not metaphysically brute (E1), then the best explanation of why we cannot explain the correlation is that there is no such correlation. Next step: if the no correlation explanation is best, then it is likely that there is no correlation. Hence (E7)—it is likely there is no correlation—that is, we are justified in thinking there is no such correlation.

But to determine whether the no reliability explanation is best, we must of course consider alternative explanations and assess their comparative explanatory merit.

Enoch does not do this, understandably, because he does not present his argument as an IBE argument. But to make such an IBE argument work, we would need to consider rival explanations and explicitly argue that they are not as good as the no reliability "eliminativist" explanation. What might such rival explanations be? It could be that the better explanation for why we do not have a plausible explanation of moral reliability (on robust moral realism) is that:

- a. We have not sufficiently looked for a plausible explanation of moral reliability,
- b. We have little to no grip on the plausible causes of our moral beliefs,
- c. We have little to no grip on the plausible content of the moral facts,
- d. We are “cognitively closed” to a plausible explanation of moral reliability. (i.e. we are *in principle* unable to obtain a plausible explanation),
- e. Our evolved brains are simply not good at resolving this sort of explanatory question,
- f. Moral reliability is a product of chance.²⁴

If we have not sufficiently looked for an explanation—in the case of robust moral realism, a “third-factor” explanation—that could perhaps best explain why we have not found one. Or it could be that we have little to no grip on the plausible causes of our moral beliefs or what the moral facts might plausibly be: if we do not have a grip on these two factors, it should be no surprise that we cannot explain why they are so

²⁴ Another possible explanation: some of our general background beliefs (e.g. in human evolution or in some of the criteria of a good explanation) that function as standards or constraints on the sorts of explanations we find plausible are false and thus the reason why we have no plausible explanation of moral reliability is because we are not countenancing explanations that fail to cohere with our (false) background beliefs. Given how widely shared our general background beliefs are and how monumental a task it would be to assess them, let us presume that they are roughly right and thus that this further explanatory possibility does not obtain and is thus not a “live” rival explanation.

correlated. Or it could be that we are “cognitively closed” to a plausible explanation of tracking—that is, for some reason unable *in principle* to obtain a plausible explanation. If that were the case, then the best explanation of our failure to currently have a plausible explanation of moral reliability would not be that there is no such reliability, but rather that it is simply inaccessible to us. If not cognitively closed, it could be that we are cognitively handicapped by our evolutionary heritage—that is, our evolved brains are just not good at explaining things like moral reliability. Finally, it could be that moral reliability is a product of chance or a metaphysically brute fact. If either of these were the case, we should not expect to have a good explanation of reliability if it exists because there is no such explanation out there to be had.

Except for the assertion that the bruteness explanation is “highly unlikely,”²⁵ Enoch says nothing that would undermine or rule out any such rival explanations of our failure to explain moral reliability. So though Enoch’s inference from E1 and E6 to E7 does not work, the IBE argument suggested above and the rival explanations identified above point the way toward the development of a new and better explanatory challenge.

As the example of MATHEMATICAL RELIABILITY suggests, we need some bridge premise(s) to secure the inference from E6 to E7. Enoch’s claim of unlikely bruteness (E1) by itself cannot secure the inference since all this claim implies is the

²⁵ Enoch (2010), p. 423.

metaphysical point that there must *be* some explanation or other of the reliability of our moral beliefs if they are in fact reliable; it does not imply the epistemological point that if our moral beliefs are reliable the explanation of their reliability must be currently known or even knowable in principle by creatures like us or that we should expect ourselves to have a plausible explanation of their reliability by now. What might successful bridge premises look like? Given the counterexample to the direct move from E1 and E6 to E7—and the prospect that any general epistemic principle grounding a move of this sort will be subject to similar counterexample—the above considerations suggest that we explicitly develop the explanatory challenge on the basis of an inference to the best explanation, where the no reliability explanation is argued to be comparatively better than all candidate rivals. In addition to the comparative explanatory assessment, additional argument would be needed concerning the *epistemic implications* of IBE in this instance. After all, even if the no reliability explanation is best, it could be the best of a “poor lot”—that is, the best of an improbable set of explanations—and so the explanatory challenge, if cast as an IBE argument, could very well be epistemically impotent. My new explanatory challenge, as we shall see in Chapters 4 and 5, fleshes out this suggested argument from IBE and argues that it possesses an important epistemic upshot.

3.4 Street's Darwinian Dilemma

Now consider Sharon Street's explanatory challenge. Street (2006) targets value realism with an explanatory challenge that she dubs the "Darwinian Dilemma." Though Street's target, like Enoch's target, is realism about value or normativity in general, rather than *moral* value or normativity in particular, for reasons of economy and continuity let us fix on the moral case. Street's target then is moral realism, the view that there are moral facts understood as response-independent natural facts or non-natural facts. Such moral facts are not constituted by nor a function of our (actual or ideal) evaluative attitudes or responses (e.g. beliefs, desires, emotions, dispositions, motivations, conventions, social practices, etc).²⁶ The scope of Street's challenge does not include response-dependent or "constructivist" or "anti-realist" views that hold the moral facts are constituted by or are a function of our evaluative attitudes.²⁷ This is dialectically important for Street because she argues her own constructivist metaethics

²⁶ Street (2006), pp. 110-111. On what Street means by evaluative attitudes:

Evaluative attitudes I understand to include states such as desires, attitudes of approval and disapproval, unreflective evaluative tendencies such as the tendency to experience X as counting in favor of or demanding Y, and consciously or unconsciously held evaluative judgements, such as judgements about what is a reason for what, about what one should or ought to do, about what is good, valuable, or worthwhile, about what is morally right or wrong, and so on. (Street 2006, p.110).

²⁷ Street (2006), pp. 110-111.

escapes the Darwinian Dilemma, while her metaethical arch rival (moral realism) falls prey to it.

Street's original (2006) formulation of her Darwinian Dilemma defies easy summary and demands an ambitious piecing together—the sustained argument for the dilemma occupies fifty-five pages in *Philosophical Studies*. Since my critical target in this chapter is confined to one central move in her argument, a thorough reconstruction of Street's dilemma is unnecessary. But to forestall suspicions of strawmanning and to get clear on the place of the targeted central move in her larger argument, I must provide at least an accurate sketch of the dilemma. This would not only be useful for my critical purposes but would be a contribution to the area given that such an accurate sketch of the dilemma has arguably not yet been provided. Extant reconstructions of Street (e.g. Copp (2008), Fitzpatrick (2008), Kramer (2009), Enoch (2010)) are very quick and gloss over central moves in the argument. My reconstruction, though not exhaustive, makes explicit central moves neglected by others. My subsequent *evaluation* of the dilemma also contributes to the discussion. Extant evaluations fix on her evolutionary causal claim (e.g. Fitzpatrick (2008), Kramer (2009)) and her claim that realists have no plausible explanation of the tracking relation between the causes of our moral beliefs and the moral facts (e.g. Copp (2008), Enoch (2010)). My evaluation fixes on another central move in her argument that should be challenged and shows that her dilemma

does not work even if we grant her evolutionary causal claim and her claim that realists have no good explanation of the tracking relation.

The best entry point into Street's intricate dilemma is provided by a couple pages of her later (2008) paper where she summarizes her dilemma before defending it against David Copp's (2008) critique. (For expository convenience, allow me to underline the portions of the passage that capture the moves in the dilemma upon which I want to focus).

...There is a striking coincidence between the normative judgments we human beings think are true, and the normative judgments that evolutionary forces pushed us in the direction of making. I claim that the realist about normativity owes us an explanation of this striking fact, but has none....

The Darwinian Dilemma begins with the observation that evolutionary forces have played a tremendous role in shaping the content of human evaluative attitudes. For example, we tend to view our survival as good, our children's lives as valuable, and the fact that someone has helped us as a reason to help that person in return. From an evolutionary point of view, these and many other of our basic evaluative tendencies are no accident....

The fact of this evolutionary influence raises a puzzle for the realist about normativity. According to the normative realist, there are normative truths that hold independently of all our evaluative attitudes. Moreover, as a purely conceptual matter, these independent normative truths might be anything. In other words, for all our bare normative concepts tell us, survival might be bad, our children's lives might be worthless, and the fact that someone has helped us might be a reason to hurt that person in return....Noting this sense in which the normative truth might be anything, and noting the role of evolutionary forces in shaping the content of our basic evaluative tendencies, we may wonder

whether there is any reason to think these forces would have led us to be capable of grasping the independent normative truth posited by the realist. More specifically, we may ask the normative realist: what is the relation, if any, between the independent normative truth, on the one hand, and the evolutionary influences that shaped our evaluative attitudes, on the other? In response, the realist may either affirm or deny a relation.

Suppose first that the realist *denies* any relation between evolutionary influences and the independent normative truth (where I'm understanding "denying a relation" to be a matter of regarding the influence of evolutionary forces on our evaluative attitudes as no better than random with respect to the truth). Given how saturated with evolutionary influence our systems of normative judgment are, denying a relation leads the realist into trouble in one of two ways. Either the realist is forced to embrace a skeptical conclusion— acknowledging that our normative judgments are in all likelihood hopelessly off track, having been fundamentally shaped in their content by forces that bear no relation to the independent normative truth— or else the realist must hold that an astonishing coincidence took place— claiming that as a matter of sheer luck, evolutionary pressures affected our evaluative attitudes in such a way that they just happened to land on or near the true normative views among all the conceptually possible ones. Both of these claims are implausible, however.

Turning to the other horn of the dilemma, suppose the realist maintains that there is a relation between the evolutionary influences on our evaluative attitudes and the independent normative truth. Importantly, the realist owes us an account of what this relation is. To insist that there is a relation while failing to offer any account of what that relation might be is no more satisfactory than positing that evolutionary forces landed us on or near the truth by sheer coincidence.

What account of the relation might the realist offer? There is one obvious candidate. According to what we may call the *tracking account*, it somehow promoted reproductive success to grasp the independent normative truth, and so creatures with an ability to do so were selected for. Unfortunately for the realist, however, the tracking account is scientifically indefensible. To explain why human beings tend to make the normative judgments we do, we do not need to suppose that these judgments are *true*, and that grasping the independent normative truth

promoted reproductive success. Rather, all we need to suppose is that making these normative judgments (or rather “proto” versions of them) got us to act in ways that tended to promote reproductive success. For example, the best explanation of why we tend to value our survival is not that it’s independently true that our survival is valuable, and that it somehow promoted reproductive success to recognize this, but rather, much more simply, that creatures who valued their survival tended to do what promoted it, and therefore left more descendants. The tracking account is untenable. Moreover, I claim, no other acceptable account of the relation is available to the realist (2008, “Reply to Copp,” pp. 207-209)

Street’s Darwinian Dilemma can be sketched as follows:

Street’s Darwinian Dilemma

S1: Assume there are realist (i.e. response independent) moral facts (for *reductio*)

S2: The content of our moral beliefs has been tremendously influenced by evolutionary causes.

S3: There is a tracking relation between the evolutionary causes of our moral beliefs and the moral facts or there is not.

NON-TRACKING HORN

S4: There is a huge universe of possible moral beliefs and possible moral facts.

S5: So, given S4, if there is not a tracking relation between the evolutionary causes of our moral beliefs and the moral facts, then most of our moral beliefs are probably false.

TRACKING HORN

S6: If there is a tracking relation, then if we have no plausible explanation of it, we are justified in thinking there is no such tracking relation.

S7: We have no plausible explanation of the alleged tracking relation.

S8: So, if there is a tracking relation, we are justified in thinking there is no such tracking relation. (S6, S7)

S9: If we are justified in thinking there is no such tracking relation, we are justified in thinking our moral beliefs are probably false.

S10: So, if there is a tracking relation, we are justified in thinking our moral beliefs are probably false. (S8, S9)

S12: So, either most of our moral beliefs are probably false or we are justified in thinking so. (S3, S5, S10).

S13: But it is implausible to suppose that most of our moral beliefs are probably false or that we are justified in thinking so.

S14: So, by *reductio*, there are no realist moral facts. (S1, S12, S13)

Street's broad argument, though not deductively tight, possesses a *reductio*-style structure. Strictly speaking, her dilemma argument functions as a subargument within the broader *reductio*, so her broad argument perhaps can be best described as the "Darwinian Reductio." The broad *reductio* assumes that there are realist moral facts, employs a dilemma subargument to derive what Street takes to be "implausible" skeptical results on each horn of the dilemma, and concludes that the initial assumption of realist moral facts that propels us toward these skeptical results is false.

Street's intended target is the realist metaphysical view that posits response-independent moral facts. Strictly speaking, her actual target is narrower than this, however, since her Darwinian Dilemma does not impugn all forms of moral realism that accept the realist metaphysical claim. For it is perfectly open to the moral realist to buy a realist moral metaphysics and yet rest content with the skeptical conclusion that, by Street's lights, the realist metaphysics implies. So Street's actual target consists of moral realisms that *reject* the skeptical result she derives from realist moral metaphysics, that is, moral realisms that make some sort of positive claim about the epistemic status of our moral judgments. Nothing is amiss here, since purely metaphysical moral realisms

possess no plausibility anyhow: the claim that there are moral facts “out there” even though we have no epistemic reason to think so is a hard sell.

Street’s argument runs this way. Assume for *reductio* that there are realist moral facts, understood as response-independent (S1). Street suggests that we couple this assumption of moral realism with the empirically supported premise that evolutionary causes have tremendously influenced the content of our moral judgments (S2). The combination of these two premises, according to Street, makes for a potent skeptical cocktail.

Street frames this skeptical cocktail as a dilemma. The central question driving the dilemma is this: are the tremendously influential evolutionary causes of our moral judgments truth-tracking or not? More precisely, but equivalently for Street, are they more likely to produce mostly true beliefs than mostly false beliefs?²⁸

The dilemma begins with (S3): either there is a tracking relation between the evolutionary causes of our moral beliefs and the realist moral facts or there is not. Consider first the non-tracking horn of the dilemma. On this horn Street relies on the claim in (S2) that evolutionary causes have tremendously influenced the content of our moral beliefs in an indirect way by virtue of their influence on our inherited evaluative

²⁸ For textual evidence that this is Street’s conception of “truth-tracking” or the “tracking relation,” see Street (2008, fn. 4).

tendencies or “instincts”²⁹—more precisely, pre-conceptual and pre-linguistic motivational precursors of our moral judgments—that in their turn give rise to our full-fledged moral beliefs. It makes sense that natural selection would strongly favor certain evaluative dispositions over others in the environment of our evolutionary ancestors, given that some dispositions would have been more likely to motivate adaptive behavior than others. So we can understand how natural selection, given the strong influence of our evaluative tendencies on which moral beliefs we happen to form, has indirectly “biased” us toward certain moral beliefs over others. That is to say, we have a plausible evolutionary story for not only why we have the capacity to form moral beliefs but also why the moral beliefs we tend to form have the content that they do.³⁰

The non-tracking horn proceeds as follows. Given the tremendous influence of evolutionary causes on the content our moral beliefs, if there is no tracking relation

²⁹ Relevant passage:

...[I]t is plausible to suppose that over the course of much of our evolutionary history, what I have been calling “more basic evaluative tendencies” were genetically heritable traits, where a basic evaluative tendency may be understood very roughly as an unreflective, non-linguistic, motivational tendency to experience something as “called for” or “demanded” in itself, or to experience one thing as “calling for” or “counting in favor of” something else. We may think of these as “proto” forms of evaluative judgement. (Street 2006, p. 119)

³⁰ Street (2006), pp. 119-120; also see pp. 113-114.

between these evolutionary causes and the moral facts—more precisely, if it is not the case that these evolutionary forces are more likely to lead to mostly true beliefs than mostly false beliefs—then given that there is a “huge universe of logically possible” moral beliefs and moral facts, it would be an “extremely unlikely” coincidence that evolutionary forces pushed us toward mostly true beliefs. The relevant passage:

Of course it's possible that as a matter of sheer chance, some large portion of our evaluative judgements ended up true, due to a happy coincidence between the realist's independent evaluative truths and the evaluative directions in which natural selection tended to push us, but this would require a fluke of luck that's not only extremely unlikely, *in view of the huge universe of logically possible evaluative judgements and truths*, but also astoundingly convenient to the realist.³¹

To get clear on the argument, we must understand how Street reasons to the extreme unlikelihood of our having mostly true moral beliefs (S5) from the premise that there is a “huge universe of logically possible” moral beliefs and moral facts (S4). The unlikelihood claim depends in part on the nonexistence of a tracking relation. For if there is no tracking relation—that is, if evolutionary causes are not more likely to lead us to mostly true beliefs than mostly false beliefs—then it is plausible that it is at least *somewhat unlikely* that they would lead us to mostly true beliefs. But how does Street derive the stronger conclusion that it is *extremely* unlikely that they would do so. To

³¹ Street (2006), p. 122, emphasis mine. Also see p. 133 for a reference back to this idea. Also see p. 208 of Street (2008).

arrive here, Street reasons rather opaquely from the claim that there is a “huge universe of logically possible evaluative judgment and truths.” Her thought appears to be in this ballpark: given all the countless sets of moral beliefs we could have arrived at, all the countless sets of moral truths that could have obtained, and all the countless possible combinations of these sets of moral beliefs and sets of moral truths that could have obtained, then if there is no tracking relation, it is *extremely* unlikely that we hit upon the right combination in the sense that evolutionary causes reliably led us to a set of mostly true beliefs. It would be something like opening a ridiculously difficult combination lock on the first try with a couple random twists of the dial. Unbelievable coincidence.

Consider now the tracking horn of the dilemma, the horn that Street thinks is the “more plausible route for the realist to take.”³² On this horn, Street relies on the claim that the realist has no plausible explanation of the tracking relation between the evolutionary forces behind our moral beliefs and the moral facts (S7). She supports her explanatory claim by critiquing the “most obvious and natural” candidate explanation of the tracking relation, an explanation that she dubs the “tracking account.” Why do our moral beliefs track the moral truth rather than not? Alternatively put: why is it no coincidence that our moral beliefs track the truth? According to the tracking account, they track the truth because having evaluative tendencies to form true (rather than false)

³² Street (2006), p. 125.

moral beliefs enabled our ancestors to survive and reproduce, or more precisely, made it more likely that our ancestors would perform adaptive behavior in their environment. Hence, those evaluative tendencies were selected for, just like our tendencies to form true (rather than false) *perceptual* beliefs (e.g. the belief that this is the edge of the cliff) tended to promote adaptive behavior (e.g. navigating cliffs without falling off) and thus were selected for.

Street contends that the tracking account, however, scores rather poorly on standard measures of explanatory merit in comparison with the same account *minus* any appeal to realist moral truths, a more parsimonious account she dubs “the adaptive link account.” This ontologically leaner account simply says that having tendencies to form the moral beliefs we do tended to promote adaptive behavior on the part of our ancestors. As Street puts it: having these evaluative tendencies “forged adaptive links between our ancestors’ circumstances and their responses to those circumstances, getting them to act, feel, and believe in ways that turned out to be reproductively advantageous.”³³ No appeal is made to response-independent moral facts and, as a result, Street contends, this account is clearer and more straightforward (why in the world is it reproductively beneficial to have tendencies to form *true* moral beliefs?), more

³³ Street (2006), p. 127

parsimonious (it does not posit moral facts or properties), more explanatorily powerful (it better explains why we tend to make some moral judgments rather than others), and is thus overall much more plausible than the tracking account. We simply do not “need to posit a role for evaluative truth,” since natural facts do all the explanatory work.³⁴ She concludes that the tracking account is “scientifically indefensible” and “untenable.”³⁵

If the tracking account cannot plausibly explain the tracking relation, it is not clear what else could explain it. Street contends, more strongly, that the tracking account is the realist’s *only* explanatory option: as she puts it, “no other acceptable account of the [tracking] relation is available to the realist.”³⁶ Since the realist’s only explanatory option is implausible, the realist has no plausible explanation of the tracking relation.

But why exactly is a tracking account the realist’s *only* explanatory option? The reason Street offers for this claim is that something about the nature of realism itself “forces” the realist to the tracking account. The relevant passage:

The reason for this stems from the very nature of realism itself. The essence of the realist position is its claim that there are evaluative truths

³⁴ Street (2006), p. 129.

³⁵ Street (2008), p. 209.

³⁶ Street (2008), p. 209.

that hold independently of all of our evaluative attitudes. But because it views these evaluative truths as ultimately independent of our evaluative attitudes, the only way for realism both to accept that those attitudes have been deeply influenced by evolutionary causes and to avoid seeing these causes as distorting is for it to claim that these causes actually in some way tracked the alleged independent truths. There is no other way to go. To abandon the tracking account—in other words, to abandon the view that selective pressures pushed us toward the acceptance of the independent evaluative truths—is just to adopt the view that selective pressures either pushed us away from or pushed us in ways that bear no relation to these evaluative truths. And to take this view is just to land oneself back in the first horn of the dilemma....Realism about value, then, has no escape: it is forced to accept either the tracking account of the relation or else the view that there is no relation at all, and both of these options are unacceptable.³⁷

Street's reasoning at this juncture is rather opaque. Puzzled, one might think that surely the realist has other *possible* explanatory options, but Street argues that the realist is "forced" to the tracking account by the "very nature of realism itself." It appears Street is equivocating in this passage, however, between the tracking *account* and the tracking *relation*.³⁸ To escape the first horn of the dilemma, the realist does need to claim, as Street says, that evolutionary causes "actually in some way tracked the

³⁷ Street (2006), pp. 134-35.

³⁸ More precisely, the equivocation appears to occur in this line of the passage: "To abandon the tracking account—in other words, to abandon the view that selective pressures pushed us toward the acceptance of the independent evaluative truths—is just to adopt the view that selective pressures either pushed us away from or pushed us in ways that bear no relation to these evaluative truths." But this is incorrect. To abandon the tracking *account* is not to abandon the tracking *relation*.

alleged independent truths," but that does not imply that the realist needs to *explain why* that tracking *relation* holds in terms of the tracking *account*. The realist may offer an entirely different explanation of the tracking relation that is compatible with the existence of response-independent moral facts and the tremendous influence of evolutionary forces on the content of our moral beliefs. For instance, the realist could opt for a Leibnizean pre-established harmony explanation where God somehow makes it the case that the evolutionary processes track the realist moral facts. David Copp (2008) and David Enoch (2010), among others, have developed alternative explanations. These explanations are *possible* explanations, even if (as I argue in Chapter 4) not ultimately plausible.

More charitably, however, we could understand Street to be suggesting that every other explanation is *implausible*, indeed more implausible than the tracking account, and thus the realist is rationally compelled or "forced" to the tracking account as the best explanatory option among a poor, improbable lot of explanations. However, she appears to give no principled argument for thinking realists cannot offer plausible explanations of the tracking relation or that the tracking account is the best that they can do. That said, perhaps we can rationally reconstruct an argument for her, i.e. piece together an argument from some critical points she makes in various parts of her central (2006) article.

At one point in the paper Street critiques what we might call the *explanation from rational reflection*, the explanation of the tracking relation which suggests that reasoning upon our evolved moral judgments can help us track the moral truth. She does this with a garbage-in-garbage-out argument. Ultimately, one can test our moral judgments “only by testing their consistency with our other evaluative judgments, combined of course with judgments about the (non-evaluative) facts.” But if so, then if there is no plausible explanation of the tracking relation between the highly influential evolutionary causes of our initial moral judgments and the realist moral facts, the process of rational reflection “has really just been a process of assessing evaluative judgments that are mostly off the mark in terms of others that are mostly off the mark.” Street concludes that “reflection of this kind isn’t going to get one any closer to evaluative truth.”³⁹

At another point in the paper, Street offers an interesting critique of *byproduct explanations* of the tracking relation, explanations that propose that our capacity to track the moral facts was not selected for but rather is an incidental byproduct of some other capacity (or set of capacities) that was selected for.⁴⁰ At another point in the paper, Street makes an interesting case against what we might call the *explanation from ethical naturalism*, an explanation on which we do not need to posit anything more than natural

³⁹ The quotes above are taken from Street (2006), p. 124.

⁴⁰ Street (2006), sect 8.

facts to explain our moral beliefs and yet we can still explain why a tracking relation holds *if* we construe the moral facts as certain purely natural facts that play a role in the best evolutionary explanation of our moral beliefs.⁴¹

Now this is not the place to reconstruct and critically evaluate her arguments against these three sorts of explanation. What is necessary for our purposes is to get clear on the *form* of argument Street offers on behalf of the claim that the realist is rationally compelled to give a tracking account of the tracking relation. And we now have the major pieces of that argument. It has the structure of a process of elimination argument: given that the tracking relation is not plausibly a coincidence,⁴² there are only four sorts of candidate explanations to which the realist can appeal in order to plausibly explain the tracking relation, namely the tracking account, the explanation from rational reflection, the byproduct explanation, and the explanation from ethical naturalism. Given that the latter three, as Street argues, are implausible, the realist is rationally compelled to the tracking account. But Street has argued that the tracking account is

⁴¹ Street (2006), sect.7.

⁴² Clarification: the claim here is not the earlier claim (on the non-tracking horn of the dilemma) that the truth of most of our moral beliefs would be an “incredible” coincidence given that non-truth tracking evolutionary forces tremendously influenced the content of our present moral beliefs. Rather, the claim here is that the *tracking relation* between the evolutionary causes of our moral beliefs and the realist moral facts is not plausibly a coincidence or the result of chance.

implausible, too, so in the final analysis the realist has no plausible explanation of the tracking relation.

Now that we understand how Street supports the crucial claim that the realist has no plausible explanation of the tracking relation, we can spell out the rest of the tracking horn as follows. If there *is* in fact a tracking relation, then given that, as Street has argued, we possess no plausible explanation of why the tracking relation holds, the claim that it does hold is “no more satisfactory than positing that evolutionary forces landed us on or near the truth by sheer coincidence.”⁴³ Street contends that the existence of such a striking tracking relation “begs for an explanation,”⁴⁴ that the realist “owes us an explanation”,⁴⁵ and suggests that the positive claim that there is such a tracking relation, in light of the fact that we have no plausible explanation of it, is dissatisfactory because *epistemically defective*. What appears to make the claim that there is a tracking relation epistemically defective—as suggested by Street’s comment that the claim is “no more satisfactory than positing that evolutionary forces landed us on or near the truth by sheer coincidence”—is that it is probably false, that is, that we are justified in thinking that there is no tracking relation (S6-S8 of the reconstruction captures this

⁴³ Street (2008), p. 209.

⁴⁴ Street (2006), p. 125.

⁴⁵ Street (2008), pp. 208-209, 211.

move). If we are justified in thinking that there is no tracking relation—that is, if we are justified in thinking evolutionary causes are not more likely to lead us to true beliefs than false beliefs—then it follows that we are justified in thinking that our moral beliefs are probably false, or at least not probably true (P9 captures the principle). So then, on Street’s tracking horn of the dilemma, the skeptical conclusion in P10 is that if there is a tracking relation, we are justified in thinking that our moral beliefs are probably false.

The broad *reductio* can now be completed. Since either there is a tracking relation or there is not, one of the two skeptical conclusions derived from the respective horns of the Darwinian Dilemma must follow, assuming the derivations are successful (S12). However, Street contends that both skeptical conclusions are each unacceptable (S13). Having arrived at an implausible skeptical result—i.e. the result that one of the two implausible skeptical conclusions are true—we have to give up at least one of the premises or inferences that pressed us to it. Street insists that the assumption of realist moral facts must go, presumably because she thinks her other premises are more plausible and her inferences are sufficiently strong. She thinks it more reasonable to give up moral realism than to stick with it and accept the skeptical upshot to which it leads us.

Her own way of avoiding the Darwinian Dilemma is to reject the assumption of realist moral facts, accept a “constructivist” view of the moral facts that holds the moral

facts are constituted by our evaluative attitudes, and grasp the tracking horn by suggesting a constructivist explanation of the tracking relation that fits nicely with the adaptive link account, all the while accepting the tremendous influence of evolutionary causes on the content of our moral beliefs. ⁴⁶

3.5 Objections to Street's Inference from No Good Explanation

The move I want to target is located on the tracking horn of Street's dilemma. It is the inference expressed by S7—the inference from the premise that we have no plausible explanation of the tracking relation to the conclusion that we are justified in

⁴⁶ Street 2006, section 10. Relevant passage:

Both accounts [the realist and anti-realist accounts] offer an explanation of why it is no coincidence that there is significant overlap between evaluative truths and the kinds of evaluative judgements that natural selection would have pushed us in the direction of. The difference is that the antirealist account of the overlap is consistent with science. Antirealism explains the overlap not with any scientific hypothesis such as the tracking account, but rather with the metaethical hypothesis that value is something that arises as a function of the evaluative attitudes of valuing creatures—attitudes the content of which happened to be shaped by natural selection. The breaking of our bones is bad, in other words, and we're well aware of this. But the explanation is not that it is true independently of our attitudes that the breaking of our bones is bad and we were selected to be able to notice this; the explanation is rather that we were selected to take the breaking of our bones to be bad, and this evaluative judgement withstands scrutiny from the standpoint of our other evaluative judgements [and thus stands in reflective equilibrium] (to speak, for example, in the voice of the constructivist antirealist). (Street 2006, p.154)

thinking there is no such tracking relation. Though different,⁴⁷ it is structurally similar to the inference targeted in Enoch's challenge.

Though the inference need not be deductively tight, it is at least unclear how the inference could be made to work. We can appreciate the rather large gap in the inference by considering a few cases that undermine the general principle that apparently underlies it, that is, the general principle that if there is no plausible explanation of the tracking relation between the causes of our beliefs and the corresponding facts, then even if there is in fact such a tracking relation, we are justified in thinking there is none. Consider this case:

ANCESTORS' PERCEPTUAL RELIABILITY: Before our ancestors had any plausible explanation of their perceptual reliability, at least some of them nonetheless justifiably believed and knew that their perceptual beliefs were reliable.

The idea is that though our ancestors could not explain *how* their perceptual beliefs were reliable, at least some of them nonetheless could presumably know *that* they were reliable from (say) sensory experience. Galileo could presumably know that his telescopic observations were reliable, even though he knew nothing about geometric

⁴⁷ Recall that Enoch's inference runs this way: if the correlation between our beliefs and the corresponding robust moral facts is not brute and we have no plausible explanation of the correlation—in the sense specified by Enoch that the standard forms of explanation (namely, causal and constitutivist explanations) are ruled out in principle and we have no plausible "third-factor" explanations—then we are justified in thinking there is no such correlation.

optics and had what we now know to be a false theory of visual perception. Today many people, and perhaps small children, though they are unable to plausibly explain their perceptual reliability, can nonetheless know their own perceptual reliability.

Such cases are plausibly actual cases, but their force as counterexamples to the principle underlying Street's inference does not hinge on their actuality. Their (realistic) possibility would be enough to throw a wrench in Street's inference. They suggest that just as we can possibly know *that* something is the case without being able to explain *how* it is the case, we can possibly know that our moral beliefs are truth-tracking or reliable without being presently able to offer a plausible explanation of that reliability. Such cases suggest that we cannot reasonably infer the nonexistence of a tracking relation *merely* from Street's claim that we have no plausible explanation of the tracking relation.

Another way to see the problem with the inference and the necessity of bridge premises is to consider the scenarios introduced in our earlier discussion of Enoch. The plausibility of any one of these scenarios would make trouble for Street's inference. For instance, it could be that there has not been a careful search for a good explanation of the tracking relation (on moral realism). If that were the case, then our present failure to obtain such an explanation of the tracking relation would not make us justified in thinking there is none. Or it could be that, as an epistemic matter, we are in the dark

when it comes to the causes behind our moral beliefs. It is empirically plausible, according to Street, that evolutionary processes have had a saturating influence on the development of our moral beliefs, but this claim has been challenged (e.g. see Fitzpatrick (2008)). If we do not have an epistemic grip on the processes behind our moral beliefs, then it is unreasonable to infer from our present inability to explain the truth-tracking character of those processes the conclusion that such processes are not truth-tracking after all.

Parallel points could be made with regard to the other scenarios mentioned earlier in our discussion of Enoch. Consideration of these scenarios suggests a way of bridging the inference between the premise that we have no plausible explanation of moral reliability and the conclusion of reliability skepticism. We could argue that since none of these scenarios plausibly obtains, the best explanation for why we have no plausible explanation of moral reliability is that there is no such reliability. This is the route taken by the new explanatory challenge that I develop in Chapters 4 and 5.

3.6 Concluding Remarks

This chapter has argued that even if we grant that moral realists possess no plausible explanation of moral reliability or moral truth-tracking, prominent explanatory challenges like Enoch's and Street's that run from this premise should not bother

realists. Why? Because these challenges share a central defect, namely they do not successfully secure the central inference to reliability skepticism from the premise that realists have no plausible explanation of moral reliability. The fact that they do not secure this inference not only demonstrates the inadequacy of prominent explanatory challenges but usefully and positively guides us to the development of a new and better explanatory challenge which secures (or at least better secures) the inference. The subsequent two chapters develop and back up this new explanatory challenge.

Chapter 4: No Good Explanation of Moral Reliability

4.1 Introduction

The present chapter argues for a central premise that figures in explanatory challenges like Street's (2006) Darwinian Dilemma, Enoch's (2010) Epistemological Challenge, and the new explanatory challenge to be developed in the next chapter. The premise is that moral realists possess no good explanation of the epistemic reliability of the processes producing our moral judgments. This premise deserves backing because moral realists, in response to the explanatory challenges of Street and Enoch, have developed sophisticated explanations of moral reliability designed to cohere with our best empirical understanding of the origins of our moral judgments. Accordingly, this chapter evaluates two prominent realist explanations—David Copp's (2008) and David Enoch's (2010)—and also a prominent and empirically-informed rationalist explanation developed by Peter Singer (1981, 1999) that could be invoked by realists to help explain moral reliability. I argue in an empirically-informed way that these prominent explanations are implausible.

Though the empirically-informed critique of prominent realist explanations motivates the premise that moral realists have no plausible explanation of moral reliability, additional support is desirable, since it is quite possible that alternative realist explanations, perhaps not yet conceived, could do better with the empirical and

philosophical resources at hand. This chapter dims the latter prospect, however, by drawing from recent work in empirical moral psychology and the evolutionary sciences to develop two empirically-based arguments for thinking not only that Copp's, Enoch's, and Singer's respective explanations do not work, but that no plausible alternative realist explanation of moral reliability is currently available, given our best empirical understanding of the evolutionary biases and nativist psychology behind our moral judgments.

By shoring up the claim that moral realists have no plausible explanation of moral reliability, this chapter strengthens explanatory challenges like Street's, Enoch's, and the new one developed in the next chapter. This is interesting because a central critique of Street's Darwinian Dilemma—for example, developed by David Copp (2008) and Clarke-Doane (2012)—is that she has not successfully shown that realists have no good explanation of moral reliability. Similarly, Enoch (2010) argues that the explanatory challenge to robust moral realism he rationally reconstructs—the Epistemological Challenge— is defective because there is in fact a plausible explanation available, namely Enoch's (2010) own evolutionary explanation. But if the present chapter successfully shows realists—including Copp (2008) and Enoch (2010)—possess no plausible explanation of moral reliability, then it neutralizes central objections to the challenges of Street and Enoch (or to amended versions thereof) and also insulates

alternative explanatory challenges from such objections. Moreover, by shoring up the claim that moral realists have no plausible explanation of moral reliability, this chapter also ups the cost of being a moral realist, given that it is a theoretical cost to a view if it is committed to a striking phenomenon (e.g. moral reliability) that it is unable to plausibly explain.

4.2 Enoch's Explanation

Start with Enoch's explanation. Officially, Enoch offers a realist explanation of *normative* reliability, which includes reliability with respect to our moral judgments, epistemic judgments, and other normative judgments. For reasons of economy and continuity, however, let us dwell on moral reliability and set aside the separate question of the plausibility of Enoch's explanation with respect to other kinds of normativity.

Enoch's explanation begins with the weak normative assumption that survival is *pro tanto* good, good to some extent:

Assume that survival or reproductive success (or whatever else evolution "aims" at) is at least somewhat good. Not, of course, that it is always good, or that its positive value is never outweighed by other considerations, or even that it is of ultimate or of intrinsic value, or anything of the sort...All I will be relying on is the assumption that survival (or whatever) is actually by-and-large better than the alternative.¹

¹ Enoch (2010), p. 430.

Enoch then invokes two sketchy evolutionary claims. First, we possess an evolved “normative governance” mechanism that motivates us to act in accordance with our moral judgments. Second, evolutionary forces have shaped the content of our moral judgments by pushing us to form adaptive moral judgments. These evolutionary claims are compatible with a variety of more precise models of how the normative governance mechanism operates and how exactly evolutionary forces have shaped the content of our moral judgments. But relying only on the two sketchy evolutionary claims and the normative assumption about survival’s *pro tanto* goodness, Enoch advances his explanation of moral reliability. He encapsulates it in this passage:

The causal influence of selective forces only directly “pushes” us in the direction of having evolutionarily beneficial beliefs, not necessarily true ones. But here as elsewhere, the two may be systematically related. For we are the kind of creatures whose actions seem to be closely related to their normative beliefs about how they should act, or how it would be good to act, or what consequences it would be good to bring about. Our mental and motivational setup seems to include a mechanism roughly like Gibbard’s (1990, Chap. 4) “normative governance” mechanism. And this completes the explanatory story needed here: Survival (or whatever) is good; so behaving in ways that promote it is (pro tanto) good; but one efficient way of pushing us in the direction of acting in those ways is by pushing us to believe that it is good to act in those ways. And in fact, as we have just seen, it is good so to act. So the normative beliefs this mechanism pushes us to have will tend to be true.²

² Enoch (2010), p. 431.

Enoch contends in this passage that given his normative assumption that survival is *pro tanto* good, given its alleged implication that actions that promote survival are also *pro tanto* good, and given our evolved normative governance mechanism, pushing us in the direction of true moral judgments would have been one efficient way for evolutionary forces to get us to perform adaptive behavior. Hence, evolutionary forces plausibly pushed us in the direction of true moral judgments. For example, disposing us to form the true (or approximately true) moral judgment that keeping one's promises or reciprocating benefits is good would have been, given our normative governance mechanism, one efficient way of motivating us to perform such adaptive behavior.

But how significant was this push toward true moral judgments (as opposed to false moral judgments)? Alternatively put: how *reliable* (or likely to lead to objectively true moral beliefs) are the evolutionary processes Enoch posits? Enoch answers: reliable enough such that when conjoined with the operation of our *reasoning mechanisms*, the entire evolution+reasoning processes are reliable enough to meet the challenge to explain moral reliability. The natural response: how reliable is reliable enough? Enoch does not specify, but presumably the evolution+reasoning processes he characterizes are robustly reliable or highly likely to lead to true moral beliefs. This is so because a high degree of reliability is that to which moral realists commit themselves: the probability

that our moral belief-forming processes deliver us with true moral beliefs must be significantly greater than a coin flip.

But must moral realists like Enoch commit themselves to the robustly reliable production of at least some of our moral beliefs? As a sociological matter, they do—they think these processes are significantly more reliable than a coin flip. But *should* they think this? What would they lose if they gave it up? It seems they should commit to the reliability claim if they want to uphold the *Epistemic Claim* that at least some of our moral judgments are epistemically justified. Otherwise, if moral realists only commit themselves to the coin flip reliability of our moral belief-forming processes, then it is at least unclear why they should ascribe *prima facie* justification to our moral judgments. The view that our moral belief-forming processes are near just as likely to lead to false beliefs as to true beliefs would plausibly defeat whatever initial justification our moral beliefs might have conferred upon them by moral phenomenology, self-evidence, or their ilk. If so, then it is difficult to see what could confer justification upon our moral judgments. Absent a successful alternative source of justification, it appears moral realists should hold that at least some of our moral judgments are reliably produced in order to secure the *Epistemic Claim* that at least some of our moral judgments are justified. And since the *Epistemic Claim* is an essential component of a plausible form of moral realism (a purely metaphysical moral realism with no positive epistemic

commitment has nil plausibility), moral realists like Enoch understandably commit themselves to the robustly reliable production of at least some of our moral beliefs.

Now that we have a grip on the degree of reliability that Enoch and other moral realists attempt to explain and need to explain, return to Enoch's suggestion that evolutionary processes are reliable enough such that when joined with the operation of our reasoning mechanisms, the entire evolution+reasoning processes are reliable enough—that is, robustly reliable or significantly more reliable than a coin flip. To explain robust moral reliability by invoking the operation of our reasoning mechanisms, Enoch recognizes that he needs to provide an explanation of the sufficient reliability of the evolution-shaped *input* judgments that are to be fed into our reasoning mechanisms. As Enoch puts it: "some reasoning (and perhaps other) mechanisms can get our normative beliefs closer to the normative truths, so long as the starting points are not too far off."³ If our moral starting points are too far off, however, then we face a garbage-in-garbage-out problem: if the input moral judgments fed into our reasoning mechanisms are unlikely to be true, then reasoning with such judgments as premises and seeking coherence among them will not sufficiently increase the probability that the resulting

³ Enoch (2010), pp. 428-429.

output moral judgments are true. That is, a reasoning-based process from unreliable input is not plausibly robustly reliable.

How do our reasoning mechanisms work to get us closer to true moral beliefs? Enoch suggests they do so by “eliminating inconsistencies, increasing overall coherence, eliminating arbitrary distinctions, drawing analogies, ruling out initially justified beliefs whose justificatory status has been defeated later on, etc.”⁴ The use of our reasoning mechanisms, as Enoch describes them, appears to amount to the use of the standard method of wide reflective equilibrium: we weigh considered moral judgments, pursue coherence among them and considered non-moral judgments and background theories, and test them by arguments for and against. We employ these mechanisms with respect to a variety of subject matters, hence their “domain-general” character. Such reasoning mechanisms, as Enoch recognizes, are *conditionally* reliable in the sense that if the processes producing the input moral judgments are sufficiently reliable, the subsequent reasoning process producing the output judgments would be even more reliable, indeed perhaps robustly reliable. Now we can put the main pieces of Enoch’s explanation together: if the evolutionary processes behind our input moral judgments are sufficiently reliable and if our conditionally reliable reasoning mechanisms operate on

⁴ Enoch (2010), p. 428.

these input judgments to yield our final moral judgments, then we can plausibly explain the reliability of the processes behind our moral judgments.

Enoch's explanation can be summed and clarified by sketching it step-wise as follows⁵:

Enoch's Evolutionary Explanation of Moral Reliability

A1: Survival is *pro-tanto* good.

A2: If something is *pro tanto* good, actions that promote that thing are *pro tanto* good.

A3: So, actions that promote survival are *pro-tanto* good. (A1, A2)

A4: We possess a normative governance mechanism, i.e. an evolved psychological mechanism that motivates us to act in accordance with our moral judgments.

A5: Given our normative governance mechanism, pushing us toward moral judgments that cohere with the *pro tanto* goodness of actions that promote survival would have been one efficient way for evolutionary processes to get us to perform such actions.

⁵ See the earlier quote in this section, taken from Enoch (2010), p. 431.

A6: So, pushing us toward *true* moral judgments would have been one efficient way for evolutionary processes to get us to perform actions that promote survival. (A3-A5)

A7: So, the evolutionary processes behind our input moral judgments plausibly pushed us toward true moral judgments. (A6)

A8: So, the evolutionary processes behind our basic moral judgments are plausibly reliable enough such that were they fed as input into conditionally reliable reasoning mechanisms, the combined evolution+reasoning processes behind our output moral judgments would be robustly reliable. (A6, A7)

A9: Our reasoning mechanisms are in fact conditionally reliable.

A10: So, the evolution+reasoning processes behind our output moral beliefs are robustly reliable. (A8, A9)

Before assessing Enoch's explanation, first consider Copp's explanation, for both explanations share important sorts of premises. Moreover, these similarities appear no accident: I will argue that any plausible explanation of moral reliability must rely upon premises of these sorts. This holds out the prospect of assessing whether moral realists

have a plausible explanation of moral reliability by assessing whether they can vindicate certain essential premises, a prospect pursued later in the chapter.

4.3 Copp's Explanation

David Copp's (2008) evolutionary explanation of moral reliability resembles Enoch's in structure. It is importantly different, however, in that though Enoch's explanation could be invoked by robust ethical non-naturalists (like Enoch), Copp's explanation relies on a specific version of ethical naturalism, namely his own.

Copp's explanation begins with the assumption that his own metaethical account—"society-centered" ethical naturalism—is correct. On his account "a basic moral proposition, such as the proposition that torture is wrong, would be true only if the moral code that would best serve the function of enabling society to meet its needs included or entailed a relevantly corresponding norm, such as a prohibition on torture."⁶ In a word: the correct moral norms are those that help a society best meet its needs. What are the needs of a (typical) society? According to Copp, these surely include "social stability, peacefulness, and cooperation."⁷ It follows on the society-centered account, according to Copp, that the correct moral norms for (typical) societies include

⁶ Copp (2008), pp. 198-199.

⁷ Copp (2008), p. 201.

“pro-social” norms that “call for kinds of behavior that promote social stability, peacefulness, and cooperation.”⁸

Copp then invokes a series of evolutionary claims to argue that biological and cultural evolutionary processes were likely to lead to the acceptance of these “pro-social,” correct moral norms. He calls his causal explanation the four-stage “adaptive link account” because he borrows the first three stages from Sharon Street’s “adaptive link account” of the evolutionary origins of our moral judgments and adds a fourth stage which he borrows from Philip Kitcher’s (2006) account of the cultural evolution of our moral judgments. Copp characterizes the four stages, roughly, as follows:

Stage 1: Natural selection endowed our human ancestors with dispositions to form pro-social evaluative attitudes that in turn, given their evolved capacity for normative governance,⁹ disposed them to perform pro-social kinds of behavior, such as “behavior that favors one’s

⁸ *Ibid.*

⁹ Copp characterizes the capacity for normative governance as “involving a tendency to accept norms that have currency in the local culture and to govern their behavior in accord with such norms.” (2008, p. 189). Kitcher (2006, p. 172) argues that this capacity was selected for because ancestors with a tendency to act on their pro-social evaluative dispositions would have been better able to establish reproductively advantageous coalitions than those without such a tendency.

offspring and relatives, that reciprocates benefits received, and that favors cooperation."¹⁰

Stage 2: The selected pro-social dispositions characterized in Stage 1 affected our less remote ancestors's conceptual development in a way that favored the pro-social content of the evaluative judgments they became disposed to make.

Stage 3: As a result of the first two stages, natural selection endowed our less remote ancestors with dispositions to subscribe to certain basic moral norms and form corresponding moral judgments.

Stage 4: As a result of the first three stages, a process of cultural evolution developed in which pro-social norms tended to survive and become widely accepted, which in turn disposed our less remote ancestors to form moral judgments that favored pro-social behavior.¹¹

¹⁰ Copp (2008), p. 189.

¹¹ For Copp's two characterizations of the four stages, see Copp (2008), pp. 187-90 and pp. 201-02.

According to Copp, then, biological and cultural evolutionary processes were likely to lead our ancestors to accept pro-social moral norms and judgments. Since such norms are correct on his society-centered account, evolutionary processes were thus likely to lead our ancestors to accept correct moral norms and judgments. But how significant of a “push” did the operative evolutionary processes give our ancestors toward the correct moral norms (as opposed to incorrect moral norms)? That is, how reliable or truth-tracking were these evolutionary processes? Copp’s answer is Enoch’s answer: reliable enough or sufficiently truth-tracking (“quasi-tracking” in Copp’s terms¹²) such that when the operation of these evolutionary processes is combined with the operation of our conditionally reliable reasoning mechanisms, the joint evolution+reasoning processes are robustly reliable, significantly more likely than a coin flip to lead to true moral beliefs.

Copp sums his explanation neatly in this passage:

[I]f a population begins by having moral beliefs with a content predicted more or less by the adaptive link account, its initial beliefs approximate sufficiently to the moral truth, by the lights of the society centered theory,

¹² In Copp’s terminology, our beliefs “quasi-track” the truth when they “tend to do *well enough* in tracking the moral truth that rational reflection can in principle correct sufficiently for any distorting influence so as to undermine the skeptical worry... For brevity, I will speak of “quasi-tracking” when I have in mind tracking to an epistemically sufficient degree.” (2008, p.194).

that, given appropriate deliberation and reflection, other things being equal, its beliefs are likely to get closer to the truth.¹³

For clarity, Copp's explanation can be usefully sketched step-wise as follows:

Copp's Evolutionary Explanation of Moral Reliability

B1: The correct moral norms are those that help a society best meet its needs (Copp's society-centered ethical naturalism).¹⁴

B2: The norms that help a society best meet its needs include "pro-social" norms that "call for kinds of behavior that promote social stability, peacefulness, and cooperation."¹⁵

B3: So, the correct norms include pro-social norms. (B1, B2)

B4: The operative evolutionary processes posited by Copp's four-stage adaptive link account were likely to lead us toward the acceptance of pro-social norms (as opposed to anti-social or socially neutral norms).

¹³ Copp (2008), pp. 200-201.

¹⁴ Copp puts it this way: "a basic moral proposition, such as the proposition that torture is wrong, would be true only if the moral code that would best serve the function of enabling society to meet its needs included or entailed a relevantly corresponding norm, such as a prohibition on torture." (2008, pp. 198-199).

¹⁵ Copp (2008), p. 201.

B5: So, the operative evolutionary processes were likely to lead us toward the acceptance of correct moral norms. (B3-B4)

B6: The operative evolutionary processes were likely to lead those who accepted moral norms to form corresponding moral judgments (i.e. judgments with the same moral content as the accepted norms).

B7: So, the operative evolutionary processes were likely to lead us toward the acceptance of true moral judgments. (B5, B6)

B8: So, the operative evolutionary processes are reliable enough such that were they conjoined with the operation of conditionally reliable reasoning mechanisms, the joint evolution+reasoning belief-forming processes would be robustly reliable. (B4-B7)

B9: Our reasoning mechanisms are conditionally reliable.

B10: So, the evolution+reasoning belief-forming processes behind our output moral judgments are robustly reliable. (B8-B9)

4.4 Clarifying the Targeted Shared Premise: Evolutionary Reliability

The step-wise reconstructions of Enoch's explanation and Copp's explanation throw into relief their structural similarity. Both explanations rely on substantive normative assumptions, evolutionary causal claims, the conditional reliability of our

coherentist reasoning mechanisms, and the epistemic claim that the evolutionary processes behind our input moral judgments are sufficiently reliable such that when conjoined with our reasoning mechanisms, the evolution+reasoning processes resulting in our output judgments are robustly reliable. Moreover, this structural similarity appears no accident: it seems that explanations of moral reliability, to be plausible, must rely on claims of these sorts. That is, plausible explanations of moral reliability apparently must rely on evolutionary causal claims, normative claims regarding substantive moral truths or the truth conditions of our moral judgments, and epistemic claims concerning the reliability of the processes behind our moral judgments. This suggests the prospect of being able to assess all candidate explanations of moral reliability by identifying and evaluating one or more of the shared claims (or sorts of claims) upon which they must rely.

Consider the claims Enoch and Copp rely upon. Though they offer some defense of their normative assumptions in their target articles or in their other work,¹⁶ their normative assumptions are highly questionable. Copp's society-centered ethical

¹⁶ Enoch (2010, pp. 432-433) for defense of his normative assumption that survival is *pro tanto* good. Copp defends his society-centered metaethical account in other work (see Copp (1995, 2007)).

naturalism, of course, possesses no short supply of critics.¹⁷ Enoch's normative assumptions, though apparently more conservative, are contentious as well. First consider Enoch's A1, the premise that survival is *pro tanto* good. If survival is *pro tanto* good—no matter whose survival it is, no matter what the conditions—it follows that the survival of an unrepentant genocidal sadist like Hitler is *pro tanto* good. But the latter is not good, arguably not even *pro tanto* good—that is, a plausible enabling condition of the *pro tanto* goodness of one's survival is that one not be an unrepentant genocidal sadist like Hitler. Hence, A1 is hard to swallow. Moreover, consider that Enoch's more precise claim in A1 is that the "aim" of evolution is *pro tanto* good—that is, reproductive fitness is *pro tanto* good, not just survival. What is problematic about this claim is that it implies that those who choose not to reproduce are missing something of moral value in their lives. It also has the disturbing implication that the propagation of a genocidal's sadist's genes is in itself *pro tanto* good.

A1 is not the only problematic normative assumption Enoch relies upon.

Consider Enoch's A2: if something is *pro tanto* good, actions that promote that thing are *pro tanto* good. The conditional is quite questionable, since even if we accepted A1 and its implication that the propagation of the genocidal sadist's genes is *pro tanto* good, it is

¹⁷ For noteworthy critiques of Copp's society-centered metaethics, see Campbell (1997) and Sobel (1998).

at least unclear whether it would follow that actions that promote the propagation of the genocidal sadist's genes are also *pro tanto* good.

As the considerations above suggest, the normative foundations that Copp and Enoch rely upon are highly contentious. But rather than take up the critical project of evaluating their normative assumptions—for instance, by challenging Copp's society-centered ethical naturalism—let us grant them for the sake of argument. For such a narrow normative critique would apparently not have interesting implications for the plausibility of explanations of moral reliability *in general*.

We should also grant Enoch's and Copp's evolutionary causal claims. For though speculative, these assumptions (or similar ones in the ballpark) appear sufficiently plausible by usual standards of evolutionary theorizing. For example, consider Enoch's evolutionary inference from A6 to A7. Just because the selection of psychological dispositions to form certain moral judgments would have been "one efficient way" to get us to perform adaptive behavior does not by itself reasonably imply that such dispositions would be selected for, since alternative psychological dispositions might be equally efficient or more efficient and thus equally likely or more likely to be selected for. For instance, a direct aversion to breaking one's promises without an accompanying disposition to form the moral judgment *that it is good to keep one's promises* might be an equally efficient or more efficient means for natural selection to

promote this kind of adaptive behavior. Nonetheless, it is plausible to suppose that dispositions to form certain moral judgments do enhance adaptive behavior.¹⁸ And evolutionary theorists typically do not demand that plausible evolutionary (i.e. selection-based) explanations, to be plausible, identify *the most efficient way* of promoting fitness. As long as the route suggested resides in the top tier of efficient ways of promoting fitness, and does not fall drastically below the other live evolutionary options, the explanation invoking such a route could very well be plausible. Those who demand more appear to demand too much from evolutionary theorizing because it is frequently too hard to identify *the most efficient route* among the evolutionary possibilities.¹⁹

Step back and consider the dialectical set up. For the sake of argument we are granting the normative and evolutionary causal claims of Enoch and Copp. What I want to challenge is their pivotal epistemic claim that the operative evolutionary processes behind our input moral judgments are sufficiently reliable. Specifically, the targeted claim is A8 in Enoch's explanation and B8 in Copp's explanation. For clarity and ease of reference, let us schematize the targeted claim as follows:

¹⁸ For empirically-informed motivation of this point, see Joyce (2007, Ch. 4).

¹⁹ For related discussion, see Stein (1996, Ch. 6).

Evolutionary Reliability: The evolutionary processes leading to our input moral judgments are sufficiently reliable such that, when combined with the operation of our conditionally reliable reasoning mechanisms, the joint evolution+reasoning processes behind our output moral judgments are robustly reliable.

A natural why-question: why challenge the shared epistemic premise of *Evolutionary Reliability*? This premise is important and philosophically interesting not only because Enoch's and Copp's explanations rely upon it but because it appears that explanations of moral reliability, to be plausible, must rely upon it. This is so for two reasons. First, it is empirically plausible to suppose that biological and cultural evolutionary processes significantly shape the content of our basic moral judgments ("moral intuitions") that function as input into our reasoning mechanisms.²⁰ So any plausible explanation of the moral reliability of the processes behind our moral judgments must explain what makes these evolutionary processes sufficiently reliability-conferring. Second, the conditional reliability of our reasoning mechanisms must play a role in the explanation, since it is highly doubtful that the operative evolutionary

²⁰ For motivation of this point see Street (2006), Copp (2008), and James (2011), among others.

processes are robustly reliable *by themselves*. This explains why realists like David Copp and David Enoch, among others, invoke our reasoning mechanisms when trying to explain moral reliability. To my knowledge, nobody claims (at least in print) that evolutionary processes can do it alone. Nonetheless, if support is needed or wanted for the point that our reasoning mechanisms must play an explanatory role and that evolutionary processes cannot do it alone, it is provided in section 4.8 of this chapter, namely by the case against the moral reliability of various biological and cultural evolutionary contributors to our moral judgments.

The question to which we are headed is whether *Evolutionary Reliability* is plausible. If *Evolutionary Reliability* is implausible, then so are all explanations of moral reliability that rely upon it. Since plausible explanations of moral reliability, to be plausible, apparently must rely on *Evolutionary Reliability*, then its implausibility would give us reason for thinking that realists have no plausible explanation of moral reliability.

But to properly evaluate *Evolutionary Reliability*, we must first make the claim more precise. Three clarificatory questions spring to mind:

- (i) What sort of epistemic reliability, more precisely, do Enoch and Copp have in mind?

- (ii) How reliable must the evolutionary processes leading to our input moral judgments be in order to be “sufficiently reliable”?
- (iii) Since reliability involves arriving at true judgments, which moral judgments do Enoch and Copp presume are true. Or: what assumptions do Enoch and Copp make about the truth conditions of our moral judgments?

Start with the first question concerning *Evolutionary Reliability*: what sort of epistemic reliability, more precisely, do Enoch and Copp have in mind? Both Enoch and Copp appear to operate with a rough probabilistic notion of process reliability or truth-tracking, as opposed to a modal notion of reliability as actual world process reliability or actual-plus-close-worlds process reliability.²¹ The basic idea of the probabilistic conception of process reliability is this:

²¹ See Enoch 2010 (p. 418, n. 12) for his operative conception of reliability, namely that a “set of beliefs is reliable only if, roughly speaking, it is not an accident that sufficiently many of them are true.” If it is not an accident that sufficiently many of our moral beliefs are true, then, roughly, they are *likely* to be true. And given that the issue Enoch is dealing with is the probability that our moral judgments are true given their origins, we can more perspicuously see his explanation of moral reliability as an explanation of probabilistic process reliability.

Copp’s (2008) paper, as a response to Street (2006), assumes Street’s probabilistic conception of truth-tracking and is rife with talk of the likelihood of processes leading us to true moral beliefs or the likelihood of our moral beliefs being true given their evolutionary origins. So we can naturally understand his explanation, too, as an explanation of

Process Reliability: process type *P* is reliable to the extent that *P* is likely to produce true beliefs *given* that *P* is instantiated.

The probability is *conditional* upon the fact that *P* is actually instantiated. Though context plausibly determines reliability assessments, in typical cases it is thought that to be robustly reliable—that is, to avoid skeptical worries regarding process *unreliability*—a process must be *significantly* more likely than not to produce true beliefs. If the chance that a moral belief-forming process results in a true moral belief resembles a coin flip, or worse, then the process is manifestly unreliable.

Consider the second question concerning *Evolutionary Reliability*: how reliable—or likely to lead to true beliefs—must these evolutionary processes behind our input moral judgments be? The answer depends, as Enoch and Copp rightly maintain, on how reliable the processes behind our input moral judgments must be in order for the operation of our conditionally reliable reasoning mechanisms to in turn enhance reliability to the requisite robust degree. But what degree of reliability this is exactly,

probabilistic process reliability rather than as an explanation of actual world reliability or actual-plus-nearby-worlds reliability.

Enoch and Copp do not address. They both suggest that the evolutionary processes behind our input moral judgments are *sufficiently* reliable. However, the indeterminacy in “sufficiently reliable” presents a problem, for the plausibility of *Evolutionary Reliability* hinges on how more precisely we are to understand this “sufficiently.” If the evolutionary processes behind our input moral judgments only have to be (say) 0.2 likely to lead to true moral beliefs in order for our reasoning mechanisms to take us the rest of the way to robust reliability, then it would be easier to support *Evolutionary Reliability* than if the operative evolutionary processes have to be more than (say) 0.5 likely to lead us to true moral beliefs.

Since it makes a big difference, which is it: nearer 0.2 or nearer 0.5? Consider again Enoch’s characterization of the operative reasoning mechanisms: these mechanisms get us closer to true moral beliefs by “eliminating inconsistencies, increasing overall coherence, eliminating arbitrary distinctions, drawing analogies, ruling out initially justified beliefs whose justificatory status has been defeated later on, etc.”²² The array of mechanisms characterized by Enoch and Copp are broadly coherentist methods of the sort employed in standard reflective equilibrium reasoning. But if the evolved input moral judgments fed into our reasoning mechanisms are highly unlikely to be true, then we encounter a garbage-in-garbage-out problem: reasoning

²² Enoch (2010), p. 428.

upon and with such highly improbable judgments will not plausibly sufficiently (if at all) increase the probability that the reasoned output moral judgments are true. So how likely to lead to true input moral judgments should the operative evolutionary processes be in order for the reasoning process from such judgments to plausibly be robustly reliable? Plausibly, near 0.5. The degree of reliability should not obviously be *more* than 0.5 in order for the operative reasoning mechanisms to in turn confer greater probability, but it should be relatively close to avoid the garbage-in-garbage-out worry.^{23 24}

Finally, consider our third question concerning *Evolutionary Reliability*: since reliability involves arriving at true judgments, which substantive moral judgments do Enoch and Copp presume are true, or what assumptions are they making about the

²³ Enoch (2010, p. 428) and Copp (2008, pp. 193-94) are sensitive to the garbage-in-garbage-out worry and suggest that the needed level of reliability is what is required to avoid the worry.

²⁴ One could respond that the degree of reliability needed to avoid the garbage-in-garbage-out worry is significantly less than 0.5, since coherentist reasoning should aim at a *wide* reflective equilibrium between our considered moral *and* non-moral judgments rather than a *narrow* reflective equilibrium between our considered moral judgments alone. That is, even if coherentist reasoning from highly unreliable moral judgments could not sufficiently increase the probability that the output reasoned moral judgments are true, coherentist reasoning from *both* highly unreliable moral judgments *and* *reliably formed non-moral judgments* could do the job. In future work I aim to engage this coherentist attempt to mitigate the garbage-in-garbage-out worry. Here let me just register my doubt that reasoning from reliable non-moral judgments (e.g. social or scientific background theories) can take us from highly unreliable moral judgments to reliable moral judgments. I can think of no plausible example where this sort of moral reasoning occurs, hence my doubt that its invocation can neutralize the garbage-in-garbage-out worry.

truth conditions of our moral judgments? To assess whether the operative evolutionary processes are reliable enough or (now we can say) at least near 0.5 likely to hit the target of true moral judgments, we need some understanding of the moral truths.

Understandably, neither Enoch nor Copp assumes a first-order normative theory and then argues that evolutionary processes were likely to lead to judgments that are true *according to that theory*. To do so would needlessly jeopardize their case, since the plausibility of their explanations would then hinge on a first-order normative theory and such theories are invariably contentious. Instead, both appear to conceive of true moral judgments as *what are ordinarily taken to be true moral judgments*. That is, they suppose that, given their respective premises, evolutionary processes are sufficiently likely to lead us to moral judgments that are ordinarily or commonsensically taken to be true.

What does the set of commonsense moral judgments look like for Enoch and Copp? Enoch includes the set of judgments that certain adaptive behaviors (e.g. keeping one's promises, reciprocating benefits) are *pro tanto* good. He also includes judgments that stand in tight coherence relations with such *pro tanto* judgments.²⁵ Though Enoch is not explicit about what the content of these latter judgments would be, presumably such

²⁵ On the explanatory role for Enoch of moral judgments that stand in tight coherence relations with the *pro tanto* judgments, see Enoch (2010), pp. 431, 435.

judgments would include commonsense *prima facie* duties regarding promising, reciprocity, beneficence, and so on.

Copp is more explicit and precise. For Copp, the set of commonsense moral judgments includes the set of “pro-social” judgments that favor social stability, peacefulness among neighbors, and cooperation. More precisely, this set of commonsense judgments includes traditional deontological judgments, harm norms that protect a wide range of entities, basic civil liberties (e.g. freedom of speech), minimal beneficence norms (e.g. a society’s members should be roughly “equally able to meet their basic needs to a decent minimal level”)²⁶ and so on. In other work Copp argues that such commonsense moral norms are exactly the moral norms that are true given the truth conditions identified by his society-centered metaethics. This is in fact one major way he supports his society-centered metaethics.²⁷

Given resolution of the three clarificatory questions above, we can now understand the claim of *Evolutionary Reliability* more precisely as follows:

²⁶ The quote is taken from Copp (1995), p. 203.

²⁷ In the target article, Copp (2008) writes: “I think it is likely that the [society-centered] theory yields a deontological moral code of a familiar kind (Copp 1995, 201–209).” (p. 200). Here he references his (1995) book, *Morality, Normativity, and Society*. In this book (pp. 201–209), Copp argues that commonsense moral norms like those mentioned in the main text of this paper are implied by his society-centered metaethics.

Evolutionary Reliability: The evolutionary processes leading to the input moral judgments to be fed into our conditionally reliable reasoning mechanisms are at least near 0.5 likely to lead us toward commonsense moral judgments.

4.5 Assessing Evolutionary Reliability on Enoch's and Copp's Explanations

Now we are well-positioned to answer the key question: is *Evolutionary Reliability* plausible? Everybody accepts that the sorts of evolutionary processes behind our moral judgments are contingent in the sense that they *could have* led us toward moral judgments that (by our current lights) are by and large not commonsensical. The question of reliability at issue, however, is toward which direction were they *likely* to lead us? (The question is one of probabilistic reliability). The answer could very well depend on which evolutionary processes we are considering. Accordingly, this section discusses the plausibility of *Evolutionary Reliability* on the sorts of evolutionary processes that Enoch posits and then its plausibility on the sorts of processes that Copp posits. The next section considers the plausibility of *Evolutionary Reliability* on *alternative* evolutionary processes that Enoch and Copp do not address, for even if the processes

invoked by Enoch and Copp fail to vindicate *Evolutionary Reliability*, other plausible processes might do the job.

Begin with Enoch's explanation. His evolutionary causal story proves skeletal. He first claims that we possess a normative governance mechanism, i.e. an evolved psychological mechanism that makes it so that we tend to act in accordance with our normative beliefs. Then he claims that pushing us toward beliefs cohering with the (*pro tanto*) goodness of survival-promoting actions would have been one efficient way for evolutionary forces to get us to perform such actions. Hence, Enoch suggests, evolutionary forces plausibly pushed us in the direction of forming such judgments. For the sake of argument, let us grant the actual operation of the broad sorts of evolutionary processes Enoch posits. The question of *Evolutionary Reliability* is this: given that the process types Enoch posits were actually instantiated and that their instances led us to our actual moral judgments, how *likely* were they to lead us toward commonsense moral judgments? To point out that they actually did, as a matter of fact, lead us toward mostly commonsensical moral judgments is not to explain the epistemic reliability of our moral judgments, for we might have been fairly lucky to land on mostly true moral beliefs: that is, our lucky landing on mostly true moral beliefs is perfectly compatible with the epistemic *unreliability* of our moral belief-forming processes. The key question of reliability is how *likely* were the evolutionary processes to lead us toward

commonsense moral judgments?²⁸ Can we sufficiently trust them or count on them to have guided us to moral truth? Two considerations suggest that, on Enoch's model, we cannot.

First consideration: the probability that the sorts of processes Enoch posits would lead us toward commonsensical judgments is opaque because Enoch has not offered much by way of a coherent and precise model of the processes. To back up the probability estimate, he owes us a thicker model of the operative processes.

Second consideration: even if Enoch thickens his model—say, by providing greater specification of the operative sorts of biological and cultural evolutionary processes—and even if we grant Enoch's causal claim in A5 that the evolutionary forces he characterizes likely pushed us toward judgments cohering with the *pro tanto* goodness of actions that promote survival, there lies a rather large gulf between this claim and the claim that the operative evolutionary forces were likely to push us toward true or commonsense moral judgments. For a great variety of conflicting yet nevertheless internally coherent systems of moral judgments—some broadly commonsensical and some downright nasty—include moral judgments cohering with

²⁸ More precisely (in type/token lingo): given that tokens of the evolutionary process types Enoch characterizes in fact led us to our actual input moral judgments, how *likely* is it that tokens of these process types would lead us to commonsense moral judgments?

the *pro tanto* goodness of actions that promote survival. So it is at least unclear that commonsensical moral systems—systems that include mostly commonsensical judgments—would likely evolve given the mere fact that evolutionary processes have tended to lead us toward judgments that cohere with the goodness of survival-promoting actions. So on Enoch’s model we cannot with much confidence expect or predict that we would wind up with a fairly commonsensical system of moral judgments.

For a case in point, consider harm norms. Even if we grant Enoch’s causal claim in A5 that the evolutionary forces he characterizes likely pushed us toward judgments cohering with the *pro tanto* goodness of actions that promote survival, this would not imply that we would likely arrive at commonsense harm norms. We might just as likely have accepted harm norms that are (by our current lights) rather morally detestable, as many of our ancestors did, such as harm norms permitting a man to beat his wife and children. We might recall detestable moral norms of our past that permitted or obligated people to enslave other people. Such norms are easily evolvable: such norms could benefit a group over its rivals by allowing it to increase its food supply, multiply its supply of laborers, and so on. Moreover, notice that such detestable moral norms are perfectly coherent with the weak normative claim that survival-promoting actions are *pro tanto* good (good *to some extent*) or survival-inhibiting actions are *pro tanto* bad (bad

to some extent). All this is to say: if all we know is that evolutionary forces were likely to push us toward moral judgments that cohere with the *pro tanto* goodness of survival-promoting actions, then for all we know such forces could have easily pushed us to detestable moral judgments. So Enoch's rather thin explanation does not plausibly explain *Evolutionary Reliability*. Since *Evolutionary Reliability* is a key component of his evolutionary explanation of moral reliability, his explanation fails the test of plausibility.

Copp's explanation, though thicker than Enoch's explanation, faces a similar problem. Consider Copp's four-stage evolutionary causal story, as laid out in Section 3, and grant the operation of the types of evolutionary processes Copp posits. Now the question of *Evolutionary Reliability* is this: given the instantiation of these types of processes, how likely were they to lead us toward commonsense moral judgments? Though Copp's model is thicker than Enoch's, it unfortunately still does not tell us much about the content of the norms we can expect to arise, other than that they are "pro-social" norms that "promote social stability, peacefulness, and cooperation" and call for "behavior that favors one's offspring and relatives, that reciprocates benefits received, and that favors cooperation" (2008, p. 189). But what would the content of these norms probably turn out to be? Does the evolutionary push toward pro-social norms help us predict the approximate content of (say) harm norms we would accept? Perhaps it plausibly rules out some candidates, since some norms are just unlikely to evolve: for

example, perhaps it predicts that we would not accept the harm norm that every person is obligated to destroy his or her own society.

But Copp's explanation does not narrow down the easily evolvable candidates to mostly commonsense norms as opposed to mostly detestable norms. For example, it is not (by our current lights) a commonsense norm that it is permissible or obligatory to kill enemy warriors, enslave their women and children, and take their land and resources, but such a nasty norm could easily emerge and be sustained on Copp's evolutionary model. After all, far from threatening social stability, the acceptance of this cooperative norm could easily be quite beneficial to a group, since by collectively acting on it a group could easily increase its food supply, expand its territory, eliminate its competition, and multiply its supply of laborers and sexual partners. A wealth of ethnographic and archaeological evidence indicates that the vast majority of primitive societies did in fact accept and act on such detestable harm norms.²⁹ All this motivates the general point that cooperative norms are often not (by our present lights) commonsensical: our human past is littered with nasty norms of cooperation that direct in-group members to cooperate together to exploit and mistreat out-group members.

²⁹ See Keeley's (1996) *War Before Civilization*, esp. Chs. 6 and 7. Also see Edgerton's (1992) *Sick Societies*.

Does Copp's model predict the content of other commonsense norms regarding social hierarchy, gender equality, sharing, beneficence, social justice, war, punishment, torture, sexual behavior, parenting, infanticide, animal welfare, and so on? It is difficult to see how. Many markedly clashing yet internally coherent sets of moral norms—some commonsensical and rather nasty—could be “pro-social” in Copp's sense and are thus, for all we know, equally likely to evolve given the tendency for pro-social norms to evolve in the way sketched by Copp's model. For example, given that we are plausibly disposed to discount (or even oppose) the well-being of those outside our inner circle of family, friends, and in-group, perhaps we should expect morally dubious harm norms and beneficence norms to evolve that discount (or even oppose) the well-being of neighboring out-group members, distant others, and nonhuman animals. Until fairly recently in Western history, these morally dubious norms were apparently the dominant norms and are still intractably dominant in many parts of the globe. Copp's account does not predict that the circle of our moral concern was likely to expand in the way that it recently has expanded. On his account, it could have just as easily expanded in a much more systematically in-group favoring way or stopped short at some place that by our current lights is not the commonsense place to stop.

Could Copp thicken his model with additional empirical claims in order to explain *Evolutionary Reliability*? Though Copp appears not to recognize this (at least in

print), the possibility is suggested by his earlier book *Morality, Normativity, and Society* (1995) where he invokes empirical claims that, if true, appear able to do the work of vindicating *Evolutionary Reliability*. Before showing how, let us first get clear on the role of these empirical claims in his previous work. In *Morality, Normativity, and Society* (1995), Copp develops his society-centered metaethical account and invokes substantive empirical claims to argue that his account does not have the nasty moral implications that some might ascribe to it. This defensive move is generated by the worry that Copp's account—in virtue of identifying correct moral norms with those norms that best help a society fulfill its needs—could easily endorse nasty norms that direct a society's members to exploit outsiders, since such nasty norms could easily help a society fulfill its needs. On the contrary, says Copp, his account implies nice and commonsensical moral norms that call for compassion and kindness toward all human beings and sentient animals. To make his case, Copp relies upon crucial empirical claims. He lays out his argument in the following passages:

...a society needs to promote conditions of social harmony and stability. This need is better served if its members show compassion and kindness in their dealings with one another than if they do not.

....

Our psychology cannot easily be tuned so that we respond with compassion and kindness only to members of our own society or only to humans....Kant claims that 'he who is cruel to

animals becomes hard also in his dealings with men.' I would add that he who is cruel to people outside his society becomes hard in his dealings with members of his society. Schopenhauer claimed that 'boundless compassion for all living beings is the firmest and surest guarantee of pure moral conduct.' I would say it supports our compassion for members of our own society. These remarks imply a number of generalizations about 'psychological boundaries.'

....

Now, a society needs its members to show compassion and kindness in their dealings with one another, but it would normally have no need to have its members show no compassion and kindness, or to show less, in their dealings with non-members. In fact, given the generalizations about psychological boundaries, it seems a society is likely to do better at meeting its needs if the societal code calls for compassion and kindness toward animals and toward humans generally than if the code explicitly limits its call for compassion to members of the society. If this is correct, there is reason to think that [on the society-centered theory] a justified moral code would call on people to show compassion and kindness to humans in general and also to sentient beings that can suffer and experience pain. (1995, p. 204-205)

Copp's argument can be reconstructed as follows:

Copp's (1995) Response to the Objection from Nasty Norms

(C1): A society needs social harmony and stability.

- (C2): A society's need for social harmony and stability is "better served if its members show compassion and kindness in their dealings with one another than if they do not" (p. 204).
- (C3): A society would "normally have no need to have its members show no compassion and kindness, or to show less, in their dealings with non-members" (p. 205).
- (C4): Our psychological disposition to be compassionate and kind to members of our society strongly depends upon our psychological disposition to be compassionate and kind to out-group members and sentient animals.
- (C5): So, from (C1)-(C4), the moral code that would best enable society to meet its needs includes norms that call for generalized compassion and kindness towards humans and sentient animals. Hence, the correct moral norms, on Copp's society-centered theory, include generalized compassion and kindness norms.

To be clear: the question at issue is not whether Copp (1995) is right about the nice rather than nasty moral implications of his society-centered theory—that is, whether generalized compassion and kindness moral norms would best help societies

meet their needs. Rather, the issue is whether the empirical claims invoked by Copp (1995) could supplement Copp's (2008) later evolutionary account of the origins of our moral norms in order to enhance its capacity to explain how we would likely wind up with commonsense moral norms, like the generalized compassion and kindness norms he mentions. If indeed Copp's (1995) empirical claims (C3) and (C4) are correct, and if Copp (2008) is right that we can expect pro-social norms to evolve in the way suggested by his four-stage evolutionary story, then it does seem likely that commonsense generalized compassion and kindness norms would evolve. But the problem with premises (C3) and (C4) is that they are empirically dubious. Copp concedes that he is "engaging in some armchair sociology and psychology,"³⁰ and though this speculative approach (given the subject matter) is not objectionable in principle, the empirical implausibility of these armchair claims, when they are used to thicken Copp's (2008) evolutionary model, makes their invocation an unpromising route to vindicating *Evolutionary Reliability*.

There is no short supply of evidence against premise (C4). Our psychology in fact *can* "easily be tuned" so that we are disposed to show more kindness and compassion to in-group members than to humans generally and sentient animals. After

³⁰ Copp (1995), p. 201.

all, this in-group favoritism and partialism in one form or another is undoubtedly the *psychological norm* of our past and down to the present day. Evolved in-group and partiality biases partly explain this selectiveness, but one's socialization does so as well: as Copp recognizes, "our tendency to respond to others with kindness and compassion...is shaped by the culture and by the expectations of our society's moral code" (1995, p. 204). Hence the apparent cultural divergence we find with regard to the more determinate content of compassion and kindness norms (e.g. with regard to their scope, exception clauses, and so on). Given our in-group and partiality biases and given the apparent cultural contingency of the content of compassion and kindness norms, it is highly doubtful that there is the sort of tight psychological connection between our evolved pro-social dispositions and dispositions to generalized compassion that would enable Copp's account to predict that we would arrive at generalized compassion norms.

Premise (C3) is also empirically dubious. Plausibly, groups of our ancestors that showed compassion and kindness primarily to in-group members or select out-groups could often (though not always) better meet their needs than if they showed equal or significant amounts of compassion to rival groups or distant strangers. And surely our

ancestors could often better meet their needs by not being so compassionate and kind to sentient nonhuman animals. The same is apparently true today.³¹

So the thickened model Copp could advance, due to its reliance upon empirically unrespectable claims, could not plausibly vindicate the prediction that we would wind up with commonsensical moral norms. The absence of predictive power is a problem for the Copp-style explanation of moral reliability because recall that he claims this:

[I]f a population begins by having moral beliefs with a content predicted more or less by the adaptive link account, its initial beliefs approximate sufficiently to the moral truth, by the lights of the society centered theory, that, given appropriate deliberation and reflection, other things being equal, its beliefs are likely to get closer to the truth.³² (emphasis mine)

Copp claims that his four-stage adaptive link account predicts, approximately at least, the commonsensical content of our moral norms and judgments. But this is just not so. Even if natural selection endowed us with the capacities and dispositions Copp outlines in his explanation, and even if the biological and cultural evolutionary processes he characterizes disposed our ancestors toward “pro-social” norms and judgments, it is a rather large step from these claims to the desired conclusion that these evolutionary

³¹ The discussion of the past few paragraphs concerns commonsense norms of beneficence. On whether Copp’s account plausibly explains the reliable production of commonsense harm norms or norms of non-maleficence, see earlier discussion on pp. 139-142 of the present chapter.

³² Copp (2008), pp. 200-201.

processes were likely to lead us toward commonsense moral norms. Hence Copp's explanation, like Enoch's explanation, is unable to plausibly explain *Evolutionary Reliability*.

4.6 Assessing Evolutionary Reliability on Rationalist Explanations

So both Enoch's and Copp's respective explanations appear too thin to vindicate *Evolutionary Reliability*. In this respect, they resemble traditional rationalist explanations of moral progress or of our increasing moral reliability, explanations that invoke our reasoning faculty or capacity—typically characterized either as an intuitionist capacity or a domain-general capacity for reasoning—as the capacity that has allowed us to achieve increasing moral knowledge.³³

But the problem with traditional rationalist explanations, if proposed to explain moral reliability, is that it is not at all clear why our reasoning capacity, naturalistically understood, is likely to lead us toward commonsense beliefs, especially in light of its evolutionary heritage. Our domain-general reasoning mechanisms, if reliable, are conditionally reliable at best—that is, their moral reliability hinges on the reliability of their input moral judgments. But could the reliability of the input moral judgments in turn be satisfactorily explained by some sort of empirically respectable intuitionist

³³ See Singer (1981) and Brink (1989, pp. 208-9).

capacity, some faculty that rationalists give the name 'Reason'? The prospect appears unpromising, since an operative capacity of this sort that allegedly reliably leads us to commonsense moral knowledge appears much more like an unilluminating black box or placeholder rather than an empirically supported mechanism of judgment-formation endowed to us by our evolutionary heritage. The hypothesis that such an operative capacity lies behind sufficiently many of the input moral judgments that are in turn fed into our conditionally reliable reasoning mechanisms also does not sit well with the moral diversity that obtains. The more empirically respectable sort of explanation is that evolutionary and socio-cultural processes have pushed us toward our initial moral judgments by endowing us with various psychological dispositions or biases which are in turn activated by variable features of our natural and social environments. We then reason from these initial judgments to other judgments by employing our domain-general reasoning mechanisms. The problem, again, is that in light of this causal history there appears to be no good explanation of the reliability of the initial moral judgments that are fed into our reasoning mechanisms.

But the move from these points to the unacceptability of rationalist explanations of moral reliability could be too quick, for the points do not impugn all rationalist explanations. For example, the points appear not to impugn Peter Singer's (1981, 1999) prominent and empirically-informed rationalist explanation of why the content of our

moral judgments has moved in certain commonsensical moral directions. Singer's explanation is philosophically interesting both because it is empirically-informed and because, unlike traditional rationalist explanations, it invokes no intuitionist capacity. So the thought at this juncture is this: perhaps Singer's account could help explain moral reliability by explaining the sufficient reliability of a set of initial moral judgments that are in turn fed into our reasoning mechanisms—that is, perhaps Singer can help explain *Evolutionary Reliability* for the moral realist.³⁴

Singer offers an empirically-informed rationalist explanation of “the expanding circle”—that is, why our circle of moral concern has expanded in the direction of assigning equal moral concern and protection to the interests of all human beings (whether in-group or out-group) and indeed all sentient animals. According to Singer, the expansionist trend is irregular and incomplete but nonetheless historically visible. Unlike traditional rationalist explanations, Singer's explanation draws from empirical work in the evolutionary sciences and cognitive dissonance theory. His explanation also invokes no intuitionist capacity. Rather, Singer sees our evolved reasoning capacity as

³⁴ It is worth observing that Singer (1981, 1999) himself does not try to explain moral reliability *per se*, but rather tries to offer an empirically-informed explanation of the expansionist trend of our moral norms. Nevertheless, the suggestion here is that his explanation could be invoked by the moral realist to explain (or help explain) moral reliability. For instance, David Copp (2008) might add Singer's explanatory components to thicken and enhance his own evolutionary explanation of moral reliability, which was assessed earlier in the present chapter.

giving rise to our capacity for cognitive dissonance—our useful affect-laden disposition to be averse to inconsistencies among our judgments (useful because it helps us avoid acting on false beliefs)—which in turn gives rise to the disposition to adopt more impartial (and commonsensical, we might add) moral beliefs. He initially developed his explanation of the expansionist trend of our moral beliefs in *The Expanding Circle* (1981). He usefully encapsulates his explanation, and endorses it once again, in a more recent (1999) reply to critics. Singer says he “persists” in having the thought that “reason can take us to a broader perspective from which the good of one being is no more significant than the similar good of other.” “To explain the...thought,” Singer writes, “I shall describe, very briefly, what I was trying to do in *The Expanding Circle*.” His explanation of the expansionist trend—how reason can push us toward expansionist moral norms and judgments—runs as follows:

...I seek to explain [altruism toward strangers] by linking it to our capacity to reason. Our reasoning capacity, which confers obvious evolutionary advantages, also brought with it the capacity for what psychologists call ‘cognitive dissonance’—a sense that something in our belief system does not quite fit. I suggest that if, within our own tribe or society, we develop standards of impartial decision-making to resolve conflicts, and come to use such standards in order to justify our conduct to other members of our own group, then this process may develop a logic of its own. We may find it hard to avoid the conclusion that others outside our society are so similar to us, in relevant respects, that we should extend some form of ethical protection to them too. As reasoning begins we can *understand* that—for example— other humans suffer as we

do when they are without food or shelter. We can see that there is something odd about being concerned to give food and shelter to members of our own group and refusing them to others who suffer in the same way. This is especially likely to give rise to cognitive dissonance if our group is large enough to include members whom we do not know, or have had nothing to do with, while the outsiders are not in any way a threat to us. Thus we may be led by reason to extend to outsiders the altruism we already feel towards family and friends. Since this tendency could not be selected against without diminishing our highly advantageous capacity to reason, it has persisted. (1999, pp. 280-81; also see p. 284-285 and Singer (1981))

Singer's explanation can be sketched step-wise as follows:

Singer's Explanation of The Expanding Circle

- (R1) Our reasoning capacity was selected for.
- (R2) Our reasoning capacity brought with it a useful capacity for cognitive dissonance—i.e. a disposition to be averse to inconsistencies in one's belief system.
- (R3) Our human ancestors developed standards of impartial decision-making (where roughly equal moral concern and protection is assigned to in-group members) to resolve conflicts among in-group members.

- (R4) Given (R1)-(R3), when in-group members are members of sufficiently large groups and when out-group members are not in any way a threat to in-group members, then moral beliefs assigning more moral concern and protection to in-group members than to out-group members will likely give rise to cognitive dissonance.
- (R5) Cognitive dissonance tends to motivate those who experience it to resolve the inconsistencies that generate it.
- (R6) So, under the conditions described in (R1)-(R4), the operation of our reasoning capacity—by generating cognitive dissonance—is likely to lead us toward moral beliefs that assign equal moral concern and protection to out-group members.

Two considerations cast doubt on the ability of this explanation to help explain *Evolutionary Reliability*. First, even if we grant that the causal contributors Singer identifies—cognitive dissonance and its elicitors—are reliability-conferring, we should be skeptical about the causal strength of these contributors, which should in turn make us skeptical about their capacity to explain the reliability of the *entire* belief-forming processes of which they are (causally weak) components. Singer claims that the circle of

altruistic moral concern has been expanding from the family and local community toward out-group members and all human beings and sentient animals. But important clarificatory questions spring to mind: *whose* circle of altruistic concern has expanded in this way, *how far* has it expanded, and *when* has it expanded? In recent history, and presently too, indefinitely many ethnic groups have not considered certain other ethnic groups to be anything close to moral equals. Ethnic hostility encoded in moral norms marks the globe. And many more people refuse to assign much, if any, moral significance to the interests of nonhuman animals. Even worse, indefinitely many groups think it not just permissible, but obligatory or praiseworthy to significantly favor the interests of the members of one's own group (or the interests of the group itself) above that of certain out-group members. These deep-seated and frequently nasty partialist patterns of moral concern are not new: they saturate the ethnographic and anthropological record as well.³⁵ These considerations should temper estimates of moral progress.

However, it appears that *some* detectible moral progress has been made—the historical record reflects some trend toward commonsensically expanding the circle,

³⁵ E.g. see Edgerton (1992).

even if the expansion is irregular and far short of equal concern.³⁶ Suppose this is so. Our best history suggests that the expansionist trend is most visible during the past four hundred years and primarily in Western Europe.³⁷ This raises the question: why did the circle expand *so late* in the game and primarily in Western Europe, when all the enabling conditions identified by Singer plausibly obtained much earlier and in nearly every part of the globe (see Singer's enabling conditions identified in (R1)-(R4))? The anticipated response from Singer's corner is that countervailing forces obtained, forces which systematically neutralized or overrode the influence of the operative causal contributors he identifies. That is why it took so long. But if countervailing forces kept the circle of moral concern from expanding for so long in so many places, this makes one suspect that the causal contributors identified by Singer are relatively weak influences, for their influence has been systematically overridden time and time again by other forces, until about four hundred years ago when they allegedly start to do explanatory work in Western Europe. But if Singer's influences are relatively weak influences, then even if we grant their reliability-conferring character, their operation does not give us good reason to think that the *entire* processes that include them as weak causal components are reliability-conferring: a process with causally weak contributors that are reliability-

³⁶ E.g. Singer (1981), Brink (1989, pp. 208-9).

³⁷ Nichols (2004), Ch. 7.

conferring could very well be an unreliable process overall, since any reliability-conferring done by weak causal contributors could easily be overridden by relatively stronger causal contributors that are reliability-detracting. All this is to say: for Singer's explanation to help explain *Evolutionary Reliability*, it must be shown not only that the contributors he identifies are reliability-conferring, but that they are sufficiently strong influences on our moral beliefs. Since the above considerations suggest they are weak influences, we should thus doubt their ability to explain *Evolutionary Reliability*.

A second consideration more seriously undermines Singer's rationalist explanation if invoked to help explain *Evolutionary Reliability*: even if Singer's causal contributors were in fact very strong influences, we should doubt their reliability-conferring character. That is, even if the contributors identified by (R1)-(R5) have significantly shaped our moral judgments, we should doubt the claim that they were likely to lead us in the direction of commonsensically expanding the circle. To see why, consider Singer's move from (R1)-(R5) to (R6). Even if (R1)-(R5) are true, we cannot reasonably infer (R6), for cognitive dissonance could be mitigated or eliminated in a great variety of ways. More precisely: the inconsistency between the moral judgment assigning equal moral importance to in-group members and the moral judgment assigning less moral importance to out-group members, and thus the cognitive dissonance to which the inconsistency allegedly gives rise, could be mitigated or

eliminated in a great variety of ways. There is no reason for thinking that cognitive dissonance would be most efficiently reduced by adopting moral beliefs that are more impartial and egalitarian. After all, the dissonance could be eliminated by adopting a further belief in the moral relevance of some natural characteristic that in-group members share that out-group members do not share. Recall the infamous discriminatory beliefs that invoke differences in intelligence, gender, race, species, and so on to justify the moral inferiority of other groups, women, and nonhuman animals. Such moral beliefs still have a grip on much of the world today.

Moreover, if the cognitive dissonance is sufficiently weak—as suggested by the point above concerning the causal weakness of Singer’s contributors—then we might very well just live with it, given the minimal psychological cost it brings. We all consciously live with some amount of cognitive dissonance. But why might we be disposed to do so in this particular case? It can be in our interest, all-things-considered, to accept a certain amount of dissonance, and our partiality and in-group biases might thus easily drive us to do so in this case. For accepted impartialist moral norms and beliefs can be quite costly, insofar as we are committed to acting on them. Such norms can easily require us to give up a lot of our resources for the sake of those outside our inner circle and can easily saddle us with costly guilt when we do not comply with them. Hence it is easy to see how our partiality and in-group biases—roughly, our

dispositions to be attracted to beliefs and actions that favor our (actual or perceived) interests and the (actual or perceived) interests of our inner circle—could easily drive us to live with some cognitive dissonance, if it means we can retain our more partialist moral norms and beliefs. To sum up: it is doubtful that the cognitive dissonance Singer describes could be expected to lead us in the direction of commonsensically expanding the circle of moral concern, as opposed to leading us in other (i.e. partialist, not-so-commonsensual) directions. Hence, Singer's rationalist explanation is not a promising explanation of *Evolutionary Reliability*.

Now consider an objection to my assessment of Singer's explanation. The objection is that even if cognitive dissonance can and has been resolved in all sorts of ways by many people in the past and present, nonetheless all the moral realist needs to plausibly explain moral reliability is that the operative cognitive dissonance described by Singer is likely to lead *at least some of us* to expansionist, commonsense moral judgments. The objection could run on the observation that some of us (perhaps many of us) are much less subject to partiality bias and in-group biases than the average person, are more likely to resolve cognitive dissonance than to consciously live with it, and possess a deep-seated belief in the moral irrelevance of natural characteristics that in-group members share that out-group members do not share (e.g. intelligence, gender, race, ethnicity, species, etc). The idea is that if such individuals experience the sort of

cognitive dissonance Singer describes, they would be much more likely to resolve the dissonance by adopting expansionist moral judgments than to live with the dissonance or resolve it by accepting the moral relevance of natural characteristics. That is, the dissonance-based processes behind expansionist moral judgments of certain individuals may plausibly be reliable. To sum up the objection: all the moral realist needs to explain is how *some* people arrive reliably at *some domain* of moral judgments, and Singer's explanation from cognitive dissonance—thickened in the way suggested above—could plausibly explain that much.

Two points suggest the realist's explanatory task is not so easy. First, observe that the vast majority of moral realists are committed to the reliable formation of more than just a few moral judgments or a highly circumscribed domain of moral judgments. And not without reason. It appears realists should hold to the reliable production of judgments that they think we are epistemically justified in holding. Otherwise, if realists only commit themselves to the coin-flip reliability (or worse) of our moral belief-forming processes, then it is at least unclear why they should ascribe *prima facie* justification to our moral judgments. The view that our moral belief-forming processes are near just as likely to lead to false beliefs as to true beliefs would plausibly defeat whatever initial justification our moral beliefs might have conferred upon them by moral phenomenology, moral seemings, or their kin. If so, then it is difficult to see what

could confer justification upon our moral judgments. Absent a successful alternative source of justification, it appears moral realists should hold that *many* of our moral judgments are reliably produced in order to secure the intuitively desirable outcome that *many* of our moral judgments are epistemically justified. A commonly held *desideratum* of a good moral theory is that it cohere with our considered judgments, and among these considered judgments typically resides the judgment that many of our moral judgments are at least *prima facie* justified. To maintain coherence with this considered judgment, moral realists understandably and rightly commit to the reliable production of many of our moral judgments. Hence it is the reliability of *many* of our moral judgments that moral realists need to explain.

Second, even if realists need only commit themselves to the weak claim that some people reliably form some moral judgments like expansionist judgments, the thickened Singer explanation from cognitive dissonance does not plausibly explain even this much. For cognitive dissonance can be resolved in any number of alternative (non-expansionist) ways, not just the two ways suggested above. Given the plethora of alternative resolutions, it is doubtful that the resolution of the dissonance by adopting expansionist moral judgments is more likely than the *disjunction* of all alternative resolutions. This is important because even if each individual alternative resolution of the dissonance is improbable by itself, if the *disjunction* of all alternative resolutions is

more probable than the expansionist resolution, then the thickened Singer explanation would not plausibly explain why the belief-forming process it sketches is likely to lead to expansionist moral judgments.

But are there live alternative ways of resolving the cognitive dissonance in question? That is, are there live ways of mitigating or eliminating the dissonance that arises from the appreciated inconsistency between moral judgments assigning equal moral importance to in-group members and moral judgments assigning less moral importance to out-group members? Yes. Many beliefs, not just beliefs about the moral relevance of natural characteristics (e.g. intelligence, gender, race, species) could mitigate or resolve the dissonance. For example, beliefs about the moral relevance of one's moral merit, social utility, social position or role, or some other in-group/out-group difference could resolve it. Moreover, we could also resolve the dissonance (as some of us do) by accepting that human beings are not moral equals in light of the fact that they share no common characteristic that could plausibly ground their moral equality. After all, the ground of moral equality or equal moral status is a formidable philosophical problem: species is not obviously a plausible moral marker and since everyone differs in the degree to which they possess potential status-grounding capacities (e.g. moral agency, rational agency), it is hard to see why we would have *equal* moral status. Importantly, the cognitive dissonance also could very well be mitigated or

resolved by simply ignoring it and turning our attention to other matters. Indeed this is how we often respond to cognitive dissonance. Moreover, with further exploration of the empirical literature on cognitive dissonance, we could no doubt come up with more “live” ways of resolving the cognitive dissonance.

Now the key question is this: is the thickened expansionist resolution to the cognitive dissonance identified by Singer more probable than the *disjunction* of all alternative resolutions (like those mentioned above)? This is doubtful, which suggests we cannot sufficiently rely on the thickened dissonance-based moral belief-forming process suggested to lead us to expansionist moral judgments as opposed to non-expansionist moral judgments. This is to say: the thickened Singer-style explanation from cognitive dissonance does not plausibly explain how some people arrive at some moral judgments in a morally reliable way. The same sorts of considerations would apply to alternative ways of thickening Singer’s explanation.

We should pause and sum up the critical discussion at this point. I have argued that Enoch’s explanation, Copp’s explanation, and prominent rationalist explanations (E.g. Singer’s) do not plausibly explain *Evolutionary Reliability*. Since plausible explanations of moral reliability apparently must rely on *Evolutionary Reliability*, I take it that I have motivated the premise that moral realists have no plausible explanation of moral reliability.

Now consider a natural objection to the critique of realist explanations so far.

The critique of Enoch, Copp, and Singer is that they do not successfully explain how the causal forces they invoke (e.g. cognitive dissonance) confer sufficient reliability upon the processes leading to our moral judgments. But this suggests, according to the objection, that the realist is committed to the strong view that *all* the forces shaping our moral judgments are reliability-conferring. The objection on behalf of the realist is that even if (say) the cognitive dissonance Singer invokes does not confer sufficient moral reliability, other forces or causal contributors may very well do so. For example, perhaps evolutionary processes have endowed us with a reliable empathetic or sympathetic capacity for arriving at true moral judgments. The fact that we have this reliable means of arriving at true moral judgments does not imply, the realist could argue, that it was regularly employed in the majority of the societies described by our anthropological and ethnographic record, nor that it was not regularly overridden by other causal forces. All the moral realist needs, rather, is that we possess this reliable capacity and that it can presently allow us (or at least some of us) to reliably arrive at commonsense moral judgments.

By way of response, it is not an assumption of the critique so far that the realist is committed to the claim that each and every causal force shaping the content of our moral judgments is reliability-conferring. Rather, to clarify, the assumption argued for

above is that the realist is committed to *Evolutionary Reliability*, the claim that the evolutionary processes behind our initial moral judgments are overall sufficiently reliable. This is perfectly compatible with the fact that *some causal components* of the operative evolutionary processes are reliability-detracting considered by themselves.

That said, the objection usefully points out that even if the operative causal forces shaping past moral judgments were morally unreliable, the operative causal forces shaping our *present* moral judgments may be sufficiently reliable. That is, perhaps evolution has given us some sufficiently reliable means of arriving at true moral judgments that we can presently employ, even if our historical ancestors have by and large not employed it. This is a live possibility, especially since, in contrast with the alleged evolved faculty of rational intuition considered earlier, it does not seem unpromising to suppose that the operation of our evolved empathetic or sympathetic capacities could lead us in the direction of commonsense moral judgments. The liveness of the possibility suggests that even though the critique so far of realist explanations of *Evolutionary Reliability* motivates the premise that realists have no plausible explanation of moral reliability, more support for the premise is desirable. For it is a live possibility that alternative realist explanations, perhaps not yet conceived, could do better with the empirical and philosophical resources at hand. The next section dims this latter prospect, however, by drawing from recent empirical moral psychology and

evolutionary sciences to develop two empirically-based arguments for thinking not only that Copp's, Enoch's, and Singer's respective explanations do not work, but that no plausible realist explanation is currently available, given our best empirical understanding of the origins of our moral judgments.

4.7 Assessing Evolutionary Reliability More Generally: Arguments from Evolutionary Biases and Moral Nativism

Even if Enoch's explanation, Copp's explanation, and prominent rationalist explanations (e.g. Singer's), as they stand, do not plausibly explain *Evolutionary Reliability*, could a thicker or alternative realist explanation do so? Could *other* evolutionary processes be plugged in to replace or supplement the processes that Enoch and Copp invoke? The question is important, for if we should also doubt (deny or suspend judgment regarding) the moral reliability of other plausible evolutionary processes, then Copp and Enoch cannot simply give up their evolutionary claims and plug in alternative evolutionary processes to explain *Evolutionary Reliability*. That is, reasoned doubt about the moral reliability of evolutionary processes *more generally* would give us reason for thinking not only that Copp's, Enoch's, and Singer's respective evolutionary explanations are implausible, but that realists have no plausible evolutionary explanation of moral reliability.

Since much hangs on the question of whether we should doubt the moral reliability of other evolutionary processes, we should address the matter. Fortunately, the question is empirically tractable. One exhaustive (and exhausting) way to address the matter is to consider one-by-one all models of the evolution of our moral norms and judgments. This would first require refining the list of potential evolutionary processes to plausible evolutionary processes, which in turn would require forays into several evolutionary sciences. With plausible candidates in hand, we would then have to evaluate their moral reliability one-by-one. Since typical evolutionary models identify only significant *components* of the processes leading to our moral norms—not the *entire* processes—we would need to assess the moral reliability of the entire evolutionary processes by assessing the reliability of each of their plausible and causally significant components.

This systematic project is neither feasible nor necessary at this juncture. Instead, a more promising way to go is to identify empirically-based reasons for doubting the moral reliability of plausible evolutionary processes in general, whatever these processes are more exactly. The basic idea behind this strategy is that we do not need to evaluate the moral reliability of each entire evolutionary process behind our moral judgments if we have reasons for doubting the reliability of such processes in general. What could such reasons be? They could be reasons for doubting the moral reliability of plausible

and causally significant *components* shared by many of the evolutionary processes behind our moral judgments. If we could show that the evolutionary processes behind our moral judgments are plausibly composed of causally significant contributors that are reliability-detracting or distorting, that would give us reason for doubting the reliability of these processes. To be sure, it does not *entail* the evolutionary processes behind our moral judgments are unreliable, for there could be some strong reliability-conferring mediating causal contributor that corrects for or overrides the distorting influence of other contributors. But absent evidence of such a mediating cause, we should go where the evidence leads. If the evidence leads us to doubt the reliability of a process's major causal contributors, and no evidence suggests the presence of reliability-conferring contributors, we should doubt whether that process is reliable.

In what follows, I develop two arguments for doubting (denying or suspending judgment regarding) the reliability of significant components of the evolutionary processes behind our moral judgments. First, I develop the argument from evolutionary biases, which reasons that the plausible and causally significant contributors that we'll call *evolutionary biases* are highly contingent with respect to the content of the moral judgments toward which they lead us and are thus highly unreliable, which in turn gives us reason for doubting the reliability of the entire evolutionary processes that include them as components. Second, I develop the argument from moral nativism,

which reasons that though one might naturally suppose that nativist models of the origins of our moral judgments can support their reliable formation, nativist models that support *Evolutionary Reliability*—simple nativist models—are empirically implausible, so they do not help. Furthermore, sophisticated nativist models that are empirically plausible—namely, Chomskian models and bias-based models—do not support *Evolutionary Reliability* but rather give us reason to doubt it, since the causal contributors they posit are morally unreliable. So then, given our best empirical understanding of the origins of our moral judgments, the argument from evolutionary biases and the argument from moral nativism give us good reasons to doubt *Evolutionary Reliability*.

Now to develop the two arguments. Start with the *Argument from Evolutionary Biases*. This argument against *Evolutionary Reliability* is that causally significant evolutionary biases behind our moral judgments are reliability-detracting. Many theorists adopt a bias-based approach to explaining the content of moral norms, that is, an approach that holds that certain evolved dispositions play a significant role in the best explanation of our moral norms.³⁸ Two kinds of evolutionary biases loom large in the relevant empirical literature: innate biases and cultural selection biases.

Understanding the nature and workings of these evolutionary biases, as well as

³⁸ E.g. Sperber (1996), Boyer (2001), Nichols (2004), Sripada (2008). See Sperber (1996, Ch. 5) and Nichols (2004) for critiques of competing approaches.

understanding the moral diversity to which they plausibly give rise, will put us in a position to assess whether they are reliability-conferring or reliability-detracting (or neutral) influences. If they are reliability-detracting, then since they are plausibly major components of the evolutionary processes behind our moral judgments, we would have (inductive) reason for doubting the reliability of the entire evolutionary belief-forming processes that include them as components.

Let “innate biases” refer to innate dispositions to accept norms with certain contents over others. Our minds are not *tabulae rasae*: rather, they come prestructured to a significant degree by our evolutionary heritage. These innate structures are characterized variously as mechanisms, modules, principles, dispositions, and so on. The basic idea of there being innate biases behind our moral judgments is that evolved mental structures make it more likely that certain moral norms and judgments rather than others will emerge and be transmitted. These structures biasing us toward certain moral norms and judgments do not have moral content themselves, but rather dispose us to accept and transmit norms with certain moral contents. Though the best interpretation of psychological innateness is still a topic of lively debate,³⁹ a customary working definition will do for our purposes: let a disposition or bias be “innate” in the

³⁹ Samuels (2007).

sense of being (near) universally possessed, developmentally invariant, and either selected for or the byproduct of other cognitive structures that were themselves selected for.

Candidate innate biases include in-group biases, partiality bias, incest bias, reciprocity bias, vengeance bias, various affective biases (e.g. sympathetic biases, disgust biases), and so on. For a case study, consider our empirically well-established innate bias against incest. It is sometimes called the “Westermarck bias” after its founder Edvard Westermarck. Westermarck (1891) suggested that, owing to the likely genetic defects of offspring produced by incest, selection pressures would favor psychological mechanisms that would reduce sex between kin. He posited innate psychological mechanisms that give rise to an aversion to sex with those individuals that manifested a reliable indicator of kinship in our environment of evolutionary adaptation, namely close childhood associates—for instance, parents, siblings, and other relatives. He then suggested that this innate aversion to incest in turn gives rise to moral *norms* prohibiting incest. That is, an aversion that was originally selected for discouraging individuals from incest comes to favor or serve as a bias for the emergence of moral norms prohibiting incest. Interestingly, later empirical evidence has largely vindicated Westermarck’s hypothesis that we and our primate ancestors indeed possess this innate incest aversion or bias. Good and mounting empirical evidence also supports the

further claim that our incest norms emerged as a byproduct of our innate incest bias.⁴⁰

For one thing, the fact that incest norms show up in every culture is plausibly explained by the fact that we possess an innate bias against incest. To be clear: what is plausibly innate is not a determinate *norm* prohibiting incest, but rather the aversion to sex with childhood associates that disposes us to accept some sort of norm prohibiting incest.

If any candidate bias is innate and reliable in the sense of likely to push us in a commonsensical moral direction, it is probably the incest bias, since incest norms show up in all cultures and such norms are ordinarily taken to be true (or at least approximately true) moral norms. So the assessment of the moral reliability of our incest bias constitutes a good test case for the moral reliability of innate biases in general.

To assess the moral reliability of our incest bias, however, we need to know more about how it plausibly works. That is, even though we have good empirical evidence that it does in fact dispose us to accept moral norms forbidding incest, an important question for the assessment of reliability is how it disposes us. Once we have a grip on how it might plausibly dispose us, then we can better assess whether it could have easily disposed us toward not-so-commonsensical incest norms or whether it was likely to

⁴⁰ See Lieberman (2008) for summary of the empirical support.

dispose us to accept the sort of incest norm that we ordinarily take to be true (or approximately true).

So how more exactly could the incest bias plausibly dispose us to accept moral norms forbidding incest. Chandra Sripada, standing on the shoulders of a number of other theorists, including Westermarck himself, sketches a plausible process involving incest as a trigger for group disgust or aversion:

Let us suppose that in some group, one person decides that he or she will have sex with a parent or sibling. Because of the existence of the Westermarck aversion, this action, once it becomes widely known, will immediately be salient to almost everyone in the group and will serve to elicit powerful aversive reactions. Furthermore, it's plausible that actions that elicit powerful aversive reactions on the part of almost everyone in the group are more likely than other sorts of actions to lead to the emergence of new moral norms that forbid the offending action.⁴¹

Once the norm prohibiting incest emerges through this sort of group disgust reaction to incest, the Westermarck aversion is also likely to sustain and facilitate the *transmission* of the norm:

It is likely that the existence of the Westermarck aversion as part of innate human psychology would enhance the 'attractiveness' of a moral norm that forbids incest, thus helping maintain the moral norm over extended stretches of time.⁴²

⁴¹ Sripada (2008), p. 336.

⁴² Sripada (2008), pp.336-37.

So then not only is the incest bias plausibly an *origination bias*—a bias that favors the emergence of norms forbidding incest—it is also a *transmission bias* that favors the transmission and maintenance of norms forbidding incest once they emerge.⁴³

This is the crucial point for our purposes: as Sripada puts it, innate biases of both general sorts “typically operate at a fairly *general* level. That is, innate biases typically favor moral norms with certain *broad* thematic contents, but they leave open the potential for substantial variability with respect to the *specific* moral rule that is favored.”⁴⁴ For instance, Sripada suggests that though his scenario suggests how a new norm prohibiting incest of *some broad sort* might plausibly emerge, the way the Westermarck mechanism works—inducing an aversion to sex with close childhood associates—suggests that the

specific content of the newly formed incest prohibition, however, won’t be so easy to predict...[D]epending on the familial structure and child rearing patterns of the local group, the incest prohibition may apply more widely than just to the nuclear family. For example, if in some community children tend to be raised in extended households with their maternal cousins, then the operation of the Westermarck mechanism will, at least typically, lead to the development of an aversion to sexual unions with one’s maternal cousins. It’s plausible that in such a community, it is more likely that a newly formed incest prohibition will forbid sexual relations between maternal cousins as part of the incest prohibition. Overall, innate biases such as the Westermarck aversion seldom

⁴³ *Ibid.*

⁴⁴ Sripada (2008), p. 337.

determine in any precise or detailed way the content of the moral norms that they serve to favor. As a result, there will inevitably be a significant role for a number of other features, including the existing cultural context and, indeed, sheer happenstance, in determining the specific content of moral norms.⁴⁵

The way the incest bias works suggests that its operation could have easily led us toward incest norms that are substantively different from the ones we currently hold. For example, since our incest aversion is sensitive to childhood association, then if we were raised with our maternal cousins, but not with our paternal cousins, then we could have easily been pushed to incest norms forbidding sex with one's maternal cousins, but permitting sex with one's paternal cousins. These latter incest norms are not commonsensical. Secondly, the unreliability of the Westermarck bias is motivated by the extent of documented variability at the level the content of incest prohibitions.⁴⁶ If our incest bias plays a central causal role in leading us toward the incest norms we accept, as it plausibly does, then the empirically-confirmed divergence of incest norms gives us *prima facie* reason for thinking the incest bias is unreliable, since it pushes us in substantively different moral directions. Finally, one significant reason why incest is wrong, commonsensically, is its consequences in terms of the likely genetic defects of

⁴⁵ Sripada (2008), pp. 337-38.

⁴⁶ See Prinz (2005), pp. 281-282.

offspring. If so, then since the incest bias does not track genetic relatedness—but rather childhood association—we should not expect the bias to lead us toward correct incest norms, since genetic relatedness and childhood association can easily come apart, especially in the varied natural and social environments of our ancestors.

The moral reliability of our empirically well-established incest bias is a good test case of the moral reliability of innate biases in general. But to make the case stronger, let us consider other empirically-supported innate biases. Innate affective biases, especially biases associated with the emotion of sympathy, are often thought to be significant contributors to the evolution of morality and are not implausibly thought to *confer* reliability upon the processes leading to our moral judgments, since they initially appear likely to lead us to commonsense moral judgments (e.g. the expansionist judgments we considered earlier in our discussion of Singer’s rationalist explanation). So we would do well to assess their moral reliability.

Frans de Waal and other empirically-informed theorists situated in the Sentimentalist tradition have long and cogently argued that our evolved sympathetic emotional dispositions are among the “building blocks” of morality, that is, they are crucial ingredients in the emergence and development our moral norms and judgments. But just how exactly do our innate sympathetic biases shape our moral norms? Shaun

Nichols (2004) provides an empirically compelling model for how this works in the case of the cultural evolution of harm norms. We would do well to consider it.

Nichols (2004) advances the “Affective Resonance” hypothesis: “Normative prohibitions against action X will be more likely to survive if action X elicits...negative affect.”⁴⁷ Nichols’s basic idea is that our emotions dispose us to accept and culturally transmit norms that resonate with them— the “affective resonance” of norms increases their cultural fitness. More precisely, Nichols argues that norms that prohibit actions that are distressing or emotionally upsetting to sympathetic creatures like ourselves would be more likely to be accepted and culturally transmitted than affectively neutral norms. Our “distress-triggering” emotions—i.e. emotions associated with sympathy and empathy⁴⁸—thus bias us in the direction of norms that prohibit actions that cause distress. For example, Nichols argues that since we are distressed by the suffering that others experience when harmed, our affective biases push us in the direction of accepting moral norms that prohibit harming others.⁴⁹

⁴⁷ Nichols (2004), p. 129.

⁴⁸ Nichols offers a more fine-grained specification of these sympathetic emotions as “personal distress,” “contagious distress,” and “concern” (See e.g. Nichols (2004), p. 62).

⁴⁹ Nichols (2004), p.128-129; also see pp. 155-156.

One prominent way Nichols motivates his Affective Resonance account is by arguing that it can explain two characteristic features of the evolution of harm norms during the past four hundred years of Western European history: first, the expansion of the scope of individuals protected by harm norms; second, the expansion of the range of harms from which members of the moral community are protected.⁵⁰ Nichols illustrates this two-dimensional expansionist evolution of harm norms by documenting harm norms in European culture that have expanded to forbid cruelty to animals and that have expanded to forbid violent forms of corporal punishment (e.g. severing limbs).⁵¹ Since this pattern of moral evolution is plausibly taken to be evolution in a commonsensical moral direction, one might naturally suspect that the affective biases Nichols posits could be reliability-conferring, indeed could be invoked as major components of a “sentimentalist” explanation of the reliability of the processes behind our moral judgments. Perhaps Copp, for example, could supplement his evolutionary explanation of moral reliability by incorporating the reliability-conferring influence of Nichols’s affective biases. Indeed this could be an excellent strategy, since Nichols’s cumulative case for the influence of affective biases on our moral judgments—unlike

⁵⁰ Nichols (2004), p. 156.

⁵¹ Nichols (2004), pp. 143-147.

Singer's case for the significant influence of his rationalist causal contributors—is empirically compelling.

The key question, then, is whether the affective biases Nichols posits are reliability-conferring. Initial appearances notwithstanding, several considerations suggest that they are not.

First, our affective biases, as sympathetic dispositions, are significantly modulated and shaped by reliability-detracting in-group biases. It is well-known that we tend to experience more sympathy and empathy for those whom we perceive to be like us.⁵² This casts suspicion on the claim that affective biases are likely to push us toward accepting commonsensical expansionist harm norms. After all, these biases could easily lead us to *narrow* our circle of moral concern, as we could easily come to perceive other individuals and groups as less like us.

Second, our affective biases can easily lead us to favor certain groups or individuals whose interests are more salient to us, when the mere salience of interests is morally irrelevant. Batson et. al. (1995) supply empirical evidence of this phenomenon. They gave a vignette to subjects about a woman named Sheri who is waiting for medical

⁵² For empirical evidence in support of this point, see Stürmer et. al. (2005). In many studies that overtly and randomly assign people to different groups, the perception of others as being in one's in-group reliably results in their better treatment. (E.g. Levine, Cassidy, Brazier, and Reicher (2002)).

treatment and asked them if they desired to put Sheri on the top of the waitlist even though others on the waitlist had greater need for medical treatment. The group that was induced to empathize with Sheri was far more likely than the control group to move her up on the list despite of the greater need of the others. Such empirical evidence suggests that our affective biases can easily lead us in partialist, non-expansionist, non-commonsensical moral directions.⁵³

Third, our affective biases are subject to proximity effects, when mere proximity is morally irrelevant. Jesse Prinz offers a compelling recent example of this phenomenon:

There was an outpouring of support for the Katrina hurricane victims in the United States in 2005, and passionate expressions of empathy for the victims is still frequently expressed in public discourse here. The death toll was 1,836. A year later, an earthquake in Java killed 5,782 people and there was little news coverage in comparison. I would venture to guess that few Americans remember the incident. Nor is there much discussion of the Indian Ocean Tsunami a year before Katrina. People recall that event, but discuss it here with less pathos than Katrina. This despite the fact that the death toll was 315,000...The best explanation is that empathy increases for those who are nearby, culturally and geographically. (2011, pp. 226-227)

⁵³ For additional empirical evidence on this point, see Kogut and Ritov (2011) on the identifiable victim effect, the tendency of individuals to favor the interests of a particular identifiable person over the interests of a larger, less identifiable group with the same interests.

Since our affective biases tend to be sensitive to cultural and geographical closeness, it appears we cannot count on them to push us toward expansionist moral norms.

Let us step back and sum up the dialectic at this point. Empirically plausible innate biases like the incest bias and the affective biases Nichols posits are significant sources of the emergence and transmission of moral norms. If so, then reasoned doubt about their reliability should lead us to doubt whether the entire evolutionary processes that include them as components are sufficiently reliable, that is, should lead us to doubt *Evolutionary Reliability*. Moreover, the discussion of these biases above gives us good inductive reason for thinking that innate biases *in general* are not going to help realists explain moral reliability. For these biases are arguably the best candidates, since other empirically plausible and causally significant biases like partiality bias, in-group biases, disgust-based biases, vengeance biases, and other nefarious biases are unlikely to do the explanatory job. Such biases could easily lead us in nasty moral directions.

Evolutionary innate biases are significant components of the causal story, but there are other components. What of the moral reliability of the *cultural* part of the causal story? This question is important given that, in light of the extent of moral diversity that obtains in otherwise similar societies, cultural processes seem to play the most salient causal role in determining the particular content of our moral norms and judgments. To obtain a good assessment of the reliability of the *entire* evolutionary

processes behind our input moral judgments, we must assess the reliability of the cultural components of the causal chain.

Cultural processes include a rich array of forces. Rather than refining the list to plausible models and considering moral reliability on each model one-by-one—an infeasible project—we would do well to identify and assess the reliability of causally significant contributors to plausible cultural processes.

Social learning (e.g. imitative learning) is widely taken to be the cornerstone of the cultural transmission of norms. So let us consider why social learning takes the course it does, specifically why individuals learn certain norms as opposed to others. A lot of cultural transmission begins at home, from caregivers to children, but as individuals mature and venture outside the home, the sources of transmission quickly multiply and branch out from all directions, especially in large-scale societies. Relatives, peer-groups, prestigious individuals, and the community at large push us in the direction of certain norms over others. Given the entangled web of influences, we cannot plausibly explain, much less predict, why individuals join some peer group over others or why they learn from some prestigious persons over others. What many theorists have plausibly claimed, however, is that various innate *cultural selection biases*—including prestige bias, conformist bias, and age and gender biases—dispose individuals

toward certain sources of influence. That is, social learning is biased in certain directions.

How does this biased social learning work? New norms undoubtedly emerge, and just as today there is a significant amount of moral disagreement *within* cultures—no culture is monolithic and without dissent—so our ancestors undoubtedly encountered diversity in behavior and moral norms and had to select which ones to adopt. That is, individuals do not plausibly select their “cultural parents” or role models through an unbiased, random selection procedure, but rather were disposed or biased to select certain models over others. These cultural selection biases are not innate dispositions to accept moral norms with certain (broad) thematic contents—as was the case with innate biases above—but rather innate dispositions to internalize the norms of prestigious individuals (prestige bias), the norms that are most widely adopted in one’s community (conformist bias), or the norms adopted by a model who is of the same gender and slightly older (age and gender biases). In other words, rather than being *content* biases that favor norms with specific content over others with different content, these selection biases are *contextual biases* that favor norms with *whatever content* happens to be favored by the present cultural context.⁵⁴ There is a wealth of empirical evidence attesting to the existence of these contextual biases. And it is not hard to see why they

⁵⁴ See Boyd and Richerson (2005), pp.155-156.

would have been adaptive and thus selected for: trial-and-error learning is biologically expensive and adaptive behaviors would likely be found among prestigious persons (hence their success) and the community at large.⁵⁵

Now that we have a handle on the nature of these components of cultural evolution and how they work, we are in a position to see why we should doubt their reliability-conferring force. For as contextual biases they are content neutral and thus allow for the development of all sorts of norms, commonsensical and not-so-commonsensical. These biases will push us in the direction of nasty norms—for example, the norm that “you should kill enemy warriors and enslave their women and children”—if these norms are adopted by prestigious warriors, the community at large, and so on. New moral norms may easily emerge and be maintained, for instance, because prestigious members of a community and the community at large endorse a moral norm and punish those who violate it.⁵⁶ And since people become prestigious for all sorts of reasons and since norms emerge in a community for all sorts of reasons, it is not difficult to see how plausible cultural processes involving prestige-based imitation, conformist pressure, and norm enforcement can easily lead us to the emergence and

⁵⁵ See Boyd and Richerson (2005) for evidence and arguments that we have these biases and that they were plausibly selected for.

⁵⁶ Boyd and Richerson (1992).

maintenance of not-so-commonsensical norms (e.g. sexual norms and norms regarding torture and human sacrifice spring to mind).

Reasoned doubt about the reliability of these evolutionary contextual biases, as well as the innate content biases discussed above, in turn generates reasoned doubt about the reliability of the *entire* sorts of evolutionary processes that include these biases as significant components. In other words: we should doubt *Evolutionary Reliability*. And if we should doubt *Evolutionary Reliability*, then since plausible realist explanations of moral reliability apparently must rely on *Evolutionary Reliability*, we should doubt the availability of a plausible realist explanation of moral reliability.

Now let us develop the second empirically-based argument for doubting *Evolutionary Reliability*, the *Argument from Moral Nativism*. Moral nativism—broadly, the idea that some aspects of our moral psychology are innate or highly developmentally invariant—might naturally be invoked as an ally of *Evolutionary Reliability*, since one might naturally suppose it to support the idea that the content of the moral judgments to which we are led is fairly developmentally invariant and hence stable. That is, moral nativism might naturally be thought to suggest that we could not have easily been led to very different (non-commonsensical) moral judgments. But nativism does not help, or so I'll argue. Nativist models of the origins of our moral norms that can support *Evolutionary Reliability* are empirically implausible. So they do not help. And nativist

models that are empirically plausible do not support *Evolutionary Reliability* but rather give us positive reason to doubt it, since the causal contributors they posit are reliability-detracting. So they do not help either.

Start with an important distinction in moral psychology between *capacity nativism* and *content nativism*.⁵⁷ Capacity moral nativisms concern the innateness of psychological capacities (e.g. theory of mind⁵⁸) associated with our capacity for moral judgment. Content moral nativisms, on the other hand, concern the nature of the innate mental structures that shape the content of our moral norms and judgments. Since reliability concerns the truth of the content of our moral norms, the only sort of nativism that could potentially help explain the reliability of our moral judgments is content nativism. So to see whether nativist models can vindicate *Evolutionary Reliability*, we must focus on content nativist models and set capacity nativisms aside. Moreover, the distinction allows us to see that even if capacity nativisms are highly plausible,⁵⁹ this does not imply that *Evolutionary Reliability* is plausible or that the content nativisms that might vindicate *Evolutionary Reliability* are plausible.

⁵⁷ Sripada (2008), pp. 321-322.

⁵⁸ The theory of mind capacity is the capacity for mental state attribution.

⁵⁹ See e.g. Hauser (2006) for argument that the capacities associated with our capacity for moral judgment are innate.

Now consider content nativist models that, if plausible, can vindicate *Evolutionary Reliability*. Sripada captures such models schematically under what he calls the “Simple Innateness Model,” which posits that at least some moral norms are innate and identifies a moral norm or rule as innate if it is developmentally invariant, “if the development of the rule does not depend on any *specific* pattern of cultural input, such that different moral rules would emerge if different patterns of cultural input obtained.” On this view, “If there were innate moral norms of this sort, then they would almost certainly be cultural universals.”⁶⁰ But as Sripada (2008), Prinz (2008), and others cogently argue, the cross-cultural variation among moral norms with regard to their specific content—including (say) their variable scope and exceptions—makes it implausible to suppose many, if any, moral norms are innate. But to support *Evolutionary Reliability*—that is, to support the claim that the nativist processes behind our moral judgments are highly noncontingent with respect to the content of the moral judgments toward which they lead us and are thus likely to lead us toward commonsensical moral judgments—a proponent of the Simple Innateness Model would need to secure the claim that many (not just one or two) moral norms are innate. Since it is doubtful that there is even *one* innate moral norm, this nativist route to the vindication of *Evolutionary Reliability* does not look promising.

⁶⁰ Sripada (2008), p. 323.

A more plausible set of nativist models is captured schematically under what Sripada calls the Chomskian “Principles and Parameters Model.” Theorists advancing models of this sort contend that humans have an innate “moral grammar,” similar to the innate linguistic grammar posited by Noam Chomsky.⁶¹ On the Chomskian model, our linguistic grammar is thought of as providing every child with an innate menu of optional grammatical rules. For example, consider the *Subject Parameter*: some natural languages like English require articulated subjects in grammatically correct sentences (“I speak Spanish”), whereas other languages like Spanish and Italian impose no such requirement (“Hablo español”, where the subject *yo* is left out). There are many other such parameters. In our first few years of life which parameters get set or which rules get fixed is determined by input from our local cultural environment. For example, one hypothesis is that the grammatical rule that permits subjectless sentences gets fixed in early Spanish and Italian learners because they grow up hearing sentences without subjects, whereas English learners grow up hearing sentences with subjects. Our innate linguistic grammar, in combination with cultural input, is thus thought to explain the grammatical diversity among language groups while at the same time explaining why

⁶¹ For a *locus classicus*, see Chomsky (1957).

that diversity is constrained in certain ways, as well as how we acquire grammatical competence so early and on the basis of apparently impoverished stimuli.

Similarly, theorists such as Gil Harman (1999), Marc Hauser (2006), and John Mikhail (2011) argue that we possess an innate *moral* grammar that restricts the range of possible moral norms for us but does not necessarily determine the specific content of the norms we adopt. This innate structure would, in combination with input from the local cultural environment, help explain how we acquire moral norms so early and on the basis of apparently impoverished stimuli, why there is so much moral diversity, and why at the same time the moral diversity is constrained in certain ways. “Every newborn child,” Hauser writes, “could build a finite but large number of moral systems. When a child builds a particular moral system, it is because the local culture has set the parameters in a particular system. If you are born in Pakistan, your parameters are set in such a way that killing women who cheat on their husbands is not only permissible but obligatory, and the responsibility of family members.”⁶² How might this work more concretely? So far, only a few innate rules of moral grammar have been proposed and motivated.⁶³ Gil Harman (1999), for example, posits a rule of our innate moral grammar that could plausibly explain the universality of harm norms and their variable content.

⁶² Hauser (2006), p. 298.

⁶³ See Hauser (2006) and Mikhail (2011).

The rule is one that forbids harm to members of a protected class C—“you should not harm members of C” — where C is a “parameter” that can be specified differently by the different moral systems of different groups.⁶⁴ This would explain why we find harm norms of this general form in every culture, but would also accommodate the diversity we find in the *specific content* of harm norms, since different moral systems specify the protected class C differently. For example, in some cultures the permissibility of killing me hinges on whether I have cheated on my spouse or whether I am a member of the in-group, female, newborn, unborn, socially useful, and so on. It can also depend on whether I have in some way or to some extent harmed or dishonored you or your family. Proponents of this model suggest that the same sort of story—cross-cultural broad moral norms varying according to their cultural elaboration—could be told regarding incest and other moral domains.

Though the model faces formidable challenges to its motivations and empirical adequacy,⁶⁵ the issue at hand fortunately does not demand that we take a stand on the model’s empirical plausibility. For even if the model is good and accurate (and is shown to be so as the research program advances), it would not support *Evolutionary Reliability*.

⁶⁴ Harman (1999).

⁶⁵ See Sripada (2008) for such challenges.

If the nativist process it characterizes were the sort of process leading to our moral norms and judgments, it could have easily led us to nasty moral norms. For example, consider Harman's innate harm principle. The nativist process proceeding from this innate principle could have easily led to norms that robustly forbid us from harming in-group members and yet liberally permit us to harm out-group members.

But to explain *Evolutionary Reliability*, could not proponents of the Principles and Parameters Model just strengthen the constraints or parameters—for example, by somehow limiting the optional protected classes that *C* can specify, so that this innate grammatical rule reliably yields commonsense harm norms (as opposed to nasty norms)? The problem with this strategy is that the more rigid we make the parameters, the less able the model is to account for moral variation. And its capacity to explain moral variation is precisely what largely motivated the model over its competitors in the first place. So this model cannot, without losing major empirical plausibility points, lend support to *Evolutionary Reliability*.

A third and quite plausible set of nativist models is captured under the umbrella model Sripada favors and dubs the "Innate Bias Model."⁶⁶ On this popular model, innate biases, as described and discussed at length earlier in this section, dispose us to

⁶⁶ See Sripada (2008, pp. 332-340), Nichols (2004), and Sperber (1996) for motivation of the bias-based model.

accept moral norms of certain broad thematic contents. For example, our innate bias against incest disposes us to accept norms forbidding incest. But the key point, as discussed above, is that innate biases do not strongly determine the *specific content* of the norms they favor. That is, they are highly contingent with respect to the content of the moral judgments toward which they lead us. Hence, on the plausible Innate Bias Model we have reason for doubting that the sorts of nativist processes leading to our moral judgments are reliable.

So given *plausible* nativist models like those under the umbrellas of the Principles and Parameters Model and the Innate Bias Model, we have reason to doubt *Evolutionary Reliability*. Not only then does a consideration of plausible versions of moral nativism not support *Evolutionary Reliability*, they actually give us reason to doubt it. And once again, if we should doubt *Evolutionary Reliability*, then since plausible evolutionary explanations of moral reliability apparently must rely on it, it appears the realist has no plausible explanation of moral reliability. The argument from moral nativism and the argument from evolutionary biases above thus strengthen the *Explanatory Challenge* by motivating its central premise that realists have no plausible explanation of moral reliability.⁶⁷

⁶⁷ A clarificatory question may linger: how does the argument from moral nativism add something new to the argument from evolutionary biases, given that our nativist moral

4.8 Concluding Remarks

A natural and reasonable way to assess whether moral realists possess a plausible explanation of moral reliability is to assess prominent explanations of moral reliability that realists have offered. This chapter did so by assessing David Copp's (2008) and David Enoch's (2010) prominent evolutionary explanations, as well as various rationalist explanations (e.g. Singer's explanation). The strategy was first to identify a crucial premise that these explanations relied upon and that, interestingly, all candidate explanations apparently must rely on, namely the premise of *Evolutionary Reliability*. Then the task was to assess that premise. It was argued that Copp's explanation, Enoch's explanation, and prominent rationalist explanations failed to plausibly explain *Evolutionary Reliability*, as they stood. The prospect that their explanations could be thickened or replaced by alternative explanations that invoke

psychology can be plausibly understood, on the Innate Biases Model, in terms of innate evolutionary biases? The argument from moral nativism adds something new because one might accept that evolutionary biases are unreliable while maintaining that nonetheless we have a reliable Chomskian moral grammar that highly constrains the content of the moral judgments we could have formed. So even if evolutionary biases can't help explain moral reliability, perhaps some sort of plausible nativist mental architecture could. The argument from moral nativism, however, is intended to dim this prospect by showing that our nativist apparatus, on plausible models of it, is not plausibly morally reliable.

different evolutionary processes was then dimmed by considerations drawn from empirical work in moral psychology and the evolutionary sciences. Two empirically-based arguments—the argument from evolutionary biases and the argument from moral nativism—were given for thinking that not only that Copp’s and Enoch’s respective explanations do not work, but that there is no plausible evolutionary explanation of moral reliability available, since the best empirical understanding of the origins of our moral judgments gives us reasons for doubting *Evolutionary Reliability*, a key premise upon which all candidate explanations apparently must rely.

By motivating doubt about *Evolutionary Reliability* and thus motivating the premise that realists have no plausible explanation of moral reliability, this chapter strengthens explanatory challenges that rely on the premise, namely the challenges of Street, Enoch, and the new challenge developed in the next chapter. It can also strengthen amended versions of these challenges that rely on the premise. Moreover, by supporting the claim that realists have no explanation of moral reliability, the chapter also ups the cost of being a moral realist, since it is a theoretical cost to a view if it is committed to a striking phenomenon (e.g. moral reliability) that it is unable to plausibly explain.⁶⁸

⁶⁸ This point deserves comment. Moral reliability (on moral realism) is a striking alleged fact that “calls out” for an explanation. Though this is not the place to develop an account of

strikingness or of the conditions under which a phenomenon merits an explanation, we can appreciate the strikingness of moral reliability—on a standard probabilistic understanding of reliability—by observing that it amounts to a counterfactually stable correlation between our moral judgments and moral facts that are fixed independently of our judgments and responses. Though an actual world correlation between a (non-gerrymandered) domain of beliefs and its corresponding realist facts may sometimes plausibly be the result of chance—some individuals or groups might just be lucky believers with respect to some domain—it is difficult to believe that a *counterfactually stable* correlation between a domain of beliefs and its corresponding realist facts is the result of chance. I can think of no plausible example of such a correlation being due to chance. This sort of correlation calls out for an explanation, if anything does, just like the strong counterfactually stable correlations between smoking and lung cancer or between high tides and the moon’s proximity to the earth. It is plausibly a theoretical cost to a view if it posits such a striking phenomenon correlation for which it has no plausible explanation. Since moral realism is pressed to posit this correlation—moral realism without commitment to moral reliability loses major plausibility points (see my earlier discussion of this point on pp. 113-114 of the present chapter)—it cannot avoid the theoretical cost by giving up the correlation.

Chapter 5: A New Explanatory Challenge

5.1 Introduction

This chapter develops a new and provocative explanatory challenge to moral realism: plausibly explain moral reliability or face moral skepticism.

5.2 Target: Moral Realism

The challenge targets *minimal moral realism*, common ground upon which more metaphysically and epistemically robust forms of moral realism build. It is characterized by four familiar claims:

Cognitivist Claim: moral propositions are capable of being true or false.

Truth Claim: at least some moral propositions are true.

Epistemic Claim: we are epistemically justified in holding at least some moral beliefs.

Metaphysical Claim: the moral facts are response-independent, i.e. not constituted by nor a function of any (actual or ideal, individual or collective) human responses and attitudes.

As the *Metaphysical Claim* suggests, my realist targets include ethical non-naturalists who hold the moral facts are irreducible to natural facts and ethical naturalists who hold the moral facts are reducible to natural facts, but not reducible to human responses and attitudes. My targets do *not* include views that reject the four claims above. Importantly, my targets do not include response-dependent or “constructivist” views that reject the *Metaphysical Claim* by holding that the moral facts are reducible to human responses and attitudes.

Clarification: my explicit target is the *Epistemic Claim* of minimal moral realism. That is, if we grant the *Cognitivist Claim*, the *Truth Claim*, and the *Metaphysical Claim*, then my challenge contends that we should give up the *Epistemic Claim*. But the question naturally rises: must realists buy the *Epistemic Claim*? Could my realist targets avoid the explanatory challenge and emerge unscathed by simply dropping the *Epistemic Claim*? This is not a live option, since the plausibility of moral realism’s defining *Metaphysical Claim* hinges on its conjunction with the *Epistemic Claim*. If we are not epistemically justified in holding any of our moral beliefs (on a realist understanding of the moral facts)—or if we should suspend judgment about the matter—then the claim that there are nevertheless realist moral facts out there is a hard sell. Nobody in the philosophical literature, to my knowledge, holds such a position: moral realists, as a

sociological matter, but at least some minimal epistemic claim. Though a purely metaphysical understanding of moral realism is coherent, it possesses no plausibility.

Further clarification: my claim is not that the package of the four claims above constitutes *the correct* understanding of moral realism and thus that my challenge targets all realist views that genuinely count as such. To my mind, there is no one correct understanding of moral realism. Rather, my challenge targets paradigmatic and prominent realist views that share these four claims as common ground. This camp of realist views, as a sociological matter, is historically and presently prominent, if not dominant. This camp is also philosophically important for these realist views possess a variety of explanatory virtues that their opponents appear to lack: for example, they are widely thought to have an easier time explaining the intuitive objective bindingness or categorical normativity of the moral facts (e.g. child torture is wrong, regardless of what anybody thinks or feels) than response-dependent views or anti-realist views.

5.3 The Explanatory Challenge

With target identified—minimal moral realists (from here on I shall drop the “minimal” qualifier)—the new explanatory challenge can be sketched as follows.

Assume the premise—extensively supported in the previous chapter—that moral realists possess no plausible explanation of the reliability of the processes behind our moral judgments, that is, no plausible explanation that coheres with our best empirical

accounts of the origins of our moral judgments. What best explains the moral realist's failure to plausibly explain moral reliability? Inference to the best explanation always involves consideration of a range of live explanations. So consider the following natural candidate explanations of the moral realist's failure to explain moral reliability:

Explanation from Causal Ignorance: The best explanation for the absence of a plausible explanation of moral reliability (on moral realism) is that we philosophical experts—i.e. smart philosophers who have tried hard to explain moral reliability¹—are not aware of the best empirical explanations of the origins of our moral beliefs. That is, we cannot plausibly explain the reliability of our moral belief-forming processes at this point because we do not have a sufficient grip on the nature of these processes. If we are effectively in the dark about these processes, then we should not expect to have an explanation of their reliability by now. We should expect an explanation only with better (e.g. more complete, more precise) understanding of these causal processes than we have now.

¹ E.g. David Copp, Peter Railton, Neil Levy, Sharon Street, Richard Joyce, David Enoch, William Fitzpatrick, Michael Huemer, and many others.

Explanation from Moral Ignorance: The best explanation for why realists cannot yet explain how the processes behind our moral judgments are morally reliable or *truth-tracking* is that we do not have much of a grip on what the moral truths might plausibly be. If we are effectively in the dark about the *target* (moral truths), then we should not expect ourselves to be able to explain why our moral belief-forming processes are reliably hitting the target.

Explanation from Insufficient Effort: The failure to explain moral reliability is best explained by the fact that we have not searched hard and long enough for a plausible explanation on the best empirical explanations of the origins of our moral judgments and on plausible understandings of the moral truths. Just give us more time! With more time and effort, we should expect moral realists (or at least someone) to come up with a plausible explanation.

Explanation from Cognitive Closure: The best explanation of our explanatory failure is that we are “cognitively closed” to plausible explanations of moral reliability—i.e. such explanations are for some

reason inaccessible in principle to creatures like us. If we are cognitively closed, we should not expect to have an explanation now because we could not access it even if it exists.

Explanation from Evolutionary Handicap: Even if we are not cognitively closed to a plausible explanation of moral reliability, the best explanation of our explanatory failure could very well be that we are cognitively handicapped by our evolutionary heritage in our pursuit of such an explanation. That is, our evolved brains, due to the way they were constructed by evolutionary forces, are simply not likely to be able to explain something like moral reliability (on moral realism).

Bruteness Explanation: We have no plausible explanation of moral reliability yet because moral reliability is a metaphysically brute fact. If moral reliability is metaphysically brute, explanatory failure should prove no surprise because there is no explanation out there to be had.

Chance Explanation: Our failure to plausibly explain moral reliability is best explained by the fact that moral reliability is due to chance.

Whatever is due to chance is unexplainable, so if moral reliability is a product of chance, there is no explanation of it to be had. Hence, it is unsurprising that we have no explanation of it.

Eliminativist Explanation: The best explanation of our failure to explain moral reliability is that there is no moral reliability.

Motivated by consideration of these candidate explanations, the *Explanatory Challenge* runs as follows:

Explanatory Challenge

Let *e* refer to the fact that the moral realist has no plausible explanation of moral reliability.

- (1) *e* obtains. I.e., the moral realist has no plausible explanation of moral reliability.
- (2) *Explanation from Causal Ignorance* is a poor explanation of *e*.
- (3) *Explanation from Moral Ignorance* is a poor explanation of *e*.
- (4) *Explanation from Insufficient Effort* is a poor explanation of *e*.
- (5) *Explanation from Cognitive Closure* is a poor explanation of *e*.

- (6) *Explanation from Evolutionary Handicap* is a poor explanation of *e*.
- (7) *Bruteness Explanation* is a poor explanation of *e*.
- (8) *Chance Explanation* is a poor explanation of *e*.
- (9) *Eliminativist Explanation* is the best explanation of *e*. I.e., the fact that there is no moral reliability would (if true) best explain why the moral realist has no plausible explanation of moral reliability.
- (10) The set of explanations considered is not a bad (highly improbable) lot of explanations.
- (11) *IBE Principle*: Given explanandum *e* and available explanations $H_1...H_n$, if hypothesis H_i would (if true) best explain *e* and is not the best of a bad lot, then we should either suspend judgment regarding H_i or accept H_i .
- (12) So, via IBE, we should either suspend judgment regarding moral reliability or accept that there is no moral reliability (on moral realism). (1, 9-11)
- (13) Moral realists have no overriding reasons to accept moral reliability.

- (14) If we should either suspend judgment regarding moral reliability or accept that there is no moral reliability (on moral realism) and if moral realists have no overriding reasons to accept moral reliability, then moral realists should either suspend judgment regarding moral reliability or accept that there is no moral reliability.
- (15) So, moral realists should either suspend judgment regarding moral reliability or accept that there is no moral reliability. (12-14)
- (16) If moral realists should either suspend judgment regarding moral reliability or accept that there is no moral reliability, then moral realists have good reason to think we are not justified in holding our moral beliefs.
- (17) So, moral realists have good reason to think we are not justified in holding our moral beliefs. (15, 16)

The challenge pivots on an inference to the best explanation. The *explanandum* or thing to be explained is the moral realist's failure to plausibly explain the process reliability of our moral judgments on our best empirical accounts of the origins of those

judgments. The best explanation of the moral realist's failure to explain reliability, according to the challenge, is that there is no such reliability. From this outcome, skeptical implications reasonably follow. Or so I will argue.

Before developing and motivating the challenge, let me first situate the present chapter in relation to the previous chapter. The alleged fact that the moral realist possesses no plausible explanation of moral reliability—premise (1) of the challenge—deserves extensive support and accordingly was extensively supported in the previous chapter. The rest of the present chapter develops the challenge on the premise that realists have no plausible explanation of moral reliability, that is, on the premise that the *explanandum* in question obtains. By the end of the present chapter, then, we shall see what skeptical outcome is at stake for the moral realist, given that she has no plausible explanation of moral reliability.

5.4 Developing the Explanatory Challenge

5.4.1 Gist of the Challenge

Now we must develop the challenge. Premises (2)-(8) function to rule out rival explanations of the fact that the moral realist has no plausible evolutionary explanation of moral reliability. These explanations were introduced in the previous section. What if these candidate explanations, and others we can think up, are all poor explanations? If

invocations of cognitive closure, evolutionary handicap, metaphysical bruteness, and chance are implausible, and if we are aware of the best causal explanations and plausible moral truths, then if we have put in a lot of effort to connect the dots and explain how such-and-such plausible processes reliably lead us to such-and-such plausible moral truths (on moral realism), then our failure to explain the alleged reliability may very well be best explained by the fact that there is no such reliability. That is the driving thought of the *Explanatory Challenge*.

For concreteness, put flesh on the challenge. Suppose, not implausibly, that among others David Copp, Peter Railton, Michael Huemer, Colin McGinn, Sharon Street, David Enoch, Richard Joyce, Neil Levy, William Fitzpatrick, Peter Carruthers, Hallvard Lillehammer, Philip Kitcher, and Derek Parfit—to mention a few prominent philosophers—have given sustained attention to considering how plausible evolutionary, socio-cultural, and psychological processes could have plausibly and reliably led us to true moral judgments (on moral realism). If they come up empty handed, or if the explanations they come up with do not inspire confidence, then the best explanation of their explanatory failure appears not to be that they do not have a sufficient grip on plausible causal processes or plausible moral truths, nor is it that they simply have not tried long and hard enough, nor is it that they are cognitively closed or cognitively handicapped by evolution, nor is it that moral reliability is just an

unexplainable brute fact or product of chance. Rather, their failure to explain reliability appears best explained by the fact that there is no such reliability.

To deflect misunderstanding, it should be pointed out that the challenge does not rely on the simple “smart person” principle that if smart people have thought long and hard about how to explain some phenomenon e , have no good explanation of e , then e does not exist or we have good reason for thinking e does not exist. The smart person principle, though perhaps effective as an epistemic rule of thumb, is subject to counterexample. Just because smart people have thought long and hard about how to explain the origin of life and have no plausible explanation of it yet, this does not imply that there is no origin of life or that we have good reason for thinking so. Similarly, just because my kids or the man on the street do not have a good explanation of their perceptual reliability does not imply they should be skeptical about the reliability of their perceptual faculties. The operative epistemic principle underlying my challenge, however, is more sophisticated: it specifies a number of *additional conditions* that must be met to trigger the inference to the best explanation. That is, whereas the simple smart person principle casted as foil has the structure “If C_1 and C_2 , then p via IBE” my more sophisticated principle has the structure “If $C_1, C_2, C_3, C_4, C_5, C_6, C_7$, and C_8 , then p via IBE.” When set out this way, it is clear that my more sophisticated epistemic principle is

not equivalent to nor grounded in the simple smart person principle, and so counterexamples to the latter do not impugn the former.

For an illustration of the operative epistemic principle in action, consider the analogous case of theistic belief, for simplicity's sake the belief that the God of classical theism exists. The simple smart person principle implies that if smart people have thought long and hard about how to explain the reliability of the processes producing theistic belief and have come up with no good explanation, then the best explanation of explanatory failure in this case is that theistic beliefs are unreliably produced. This is a poor inference, since live alternative explanations first need to be dealt with—for example, the theist could explain our failure to explain the reliability of theistic belief-forming processes by pointing out that we have no grip on the nature of these processes. In contrast, my more sophisticated principle implies that if we are sufficiently aware of the broad sorts of processes behind our theistic beliefs, understand the content of alleged theistic truths, have thought long and hard about how to explain how such-and-such processes reliably lead to belief in such-and-such alleged theistic truths, then if hypotheses that invoke cognitive closure, evolutionary handicap, metaphysical bruteness, and chance are implausible, the best explanation of our failure to explain the reliability of the processes behind our theistic beliefs is that there is no such reliability. This inference is much more plausible.

Now that we understand the gist of the argument and have deflected a potential misunderstanding, consider the central question of whether rival explanations are in fact poor explanations—that is, whether premises (2)-(8) are plausible.

5.4.2 Against Causal Ignorance, Moral Ignorance, and Insufficient Effort

Sundry relevantly-informed philosophers (whom we are calling “philosophical experts”) have done their best to explain moral reliability and have done so, moreover, in light of increasing awareness of plausible explanations of the origins of our moral judgments and plausible understandings of the moral truths. Importantly, this is not to say that the relevant philosophical experts are aware of *the* correct and complete causal explanation of our moral judgments or *the* correct set of moral truths, nor is it to say that there is no reasonable disagreement among psychologists and philosophical experts, respectively. It is rather to say that philosophical experts possess a working familiarity, at least in broad outline, with plausible causal explanations and plausible moral truths. The relevant philosophical literature suggests as much.²

² A very small sample from the past three decades: Singer (1981), Railton (1985), Ruse (1986), McGinn (1993), Joyce (2001, 2006), Nichols (2004), Street (2006), Copp (2008), Fitzpatrick (2008), Enoch (2010), James (2011). Many other relevant works and literatures could be cited—for example, from the literature on moral realism, moral skepticism, evolutionary ethics, and experimental philosophy. Earlier (pre-1980) literature could be cited, too.

Sundry philosophers have seriously considered sophisticated explanations of moral reliability that invoke evolutionary processes, socio-cultural processes, proximate psychological causes, our methods of reasoning (e.g. reflective equilibrium), and explanations that combine these various components to explain moral reliability. The very sophistication, ingenuity, and empirically-informed character of these considered explanations suggest that smart moral philosophers have thought long and hard about how to explain moral reliability in light of the best explanations of our moral judgments and plausible views of the moral truths. So premises (2)-(4) seem reasonable. (See the next section for an extended reply to an anticipated objection to premise (2), namely the objection that our best empirical psychology is in poor epistemic standing and thus the best explanation of our failure to explain moral reliability (on moral realism) is the *Explanation from Causal Ignorance*).

An anticipated objection to premise (4)—that is, a point in favor of the *Explanation from Insufficient Effort*— is that religious baggage in the form of belief in god(s) has stymied moral philosophy, such that only recently has secular moral theorizing received the systematic attention it deserves. The idea is that we have no plausible explanation of moral reliability yet because we have not had enough time— free of religious baggage—to put in the effort required to come up with such an explanation. Peter Singer observes: “It is only since about 1960 that many people have

systematically studied non-religious ethics.”³ Given its recent start, it should be no surprise that secular moral philosophy is empty-handed at this point.

But Singer’s historical claim is patently false. Of the many counterexamples, the most glaring is the Utilitarian tradition starting from Bentham and Mill, which is self-consciously and thoroughly secular in every interesting sense. Derek Parfit’s weaker claim, on the other hand, is more plausible: Parfit claims that systematic secular moral philosophy is at most 150 years old.⁴ Setting Singer’s claim aside, then, would the objection from infancy be compelling on Parfit’s more plausible claim? It is difficult to see why. Consider: though we have only had a secular grip on the broad evolutionary processes behind our perceptual and mathematical beliefs for the past 150 years or so (since Darwin), we nevertheless have obtained plausible explanations of our perceptual and mathematical reliability (on realist understandings of the perceptual and mathematical facts). With a broad evolutionary understanding of our psychology in place, a plausible broad explanation of reliability naturally suggested itself. The basic idea has been that we are evolutionarily disposed to form true basic perceptual and mathematical beliefs because tending to form false perceptual and mathematical beliefs would be disastrous for our survival. For example, we evolved to believe that a cliff is in

³ Singer (1995, pp. 14-15). Cf. Parfit (1984), pp. 453-54.

⁴ Parfit (1984), pp. 453-54.

front of us and that $1+1=2$ because our ancestors and their cognitive equipment would not have survived if they had not detected the cliff or instead believed that $1+1=0$.⁵

Given that we developed broad plausible explanations of perceptual and mathematical reliability once we obtained a broad Darwinian understanding of the causal processes behind our perceptual and mathematical judgments, how much more time do moral realists need to come up with a good explanation of moral reliability before we start doubting its existence? The cases of perceptual reliability and mathematical reliability—in virtue of being fairly analogous to the case of moral reliability—suggest that it is the objector’s burden to show why we should suppose that with more time and effort secular moral philosophers will likely come up with a good explanation of moral reliability. Of course, it is *possible* that with more time and effort realists will come up with such an explanation. But the question at this point is whether it is *likely* to happen. Support for such an optimistic probability estimate appears unavailable. Absent such support, the objection from infancy does not get off the ground and thus the *Explanation from Insufficient Effort* remains a poor explanation.

As an objection to the burden-shifting argument from analogy, however, one could try to identify a relevant disanalogy between moral reliability, on the one hand,

⁵ On the evolutionary explanation of mathematical reliability, see Sosa (2002), Joyce (2006, p. 182), and De Cruz (2007).

and perceptual and mathematical reliability on the other hand that explains why it is not surprising at all that we do not have a plausible explanation yet of moral reliability even though we have had good explanations of perceptual and mathematical reliability for a long time now. One might argue, for example, that we should not be surprised that we have made much more progress in explaining perceptual and mathematical reliability than moral reliability because perceptual psychology and mathematics are based fundamentally on empirical observation and proof, respectively, whereas moral philosophy is based fundamentally on something more tenuous like moral intuitions. So the fact that we have had plausible explanations of perceptual and mathematical reliability for a long time and the fact that we do not yet have an explanation of moral reliability do not give us any reason to doubt (deny or suspend judgment regarding) the claim that we will likely come up with a plausible explanation of moral reliability.

However, it is unclear how the methodological claim regarding empirical observation and proof versus moral intuition, even if correct, could defuse the burden-shifting argument. The objection from the methodological claim appears best construed this way: owing to respective methodological differences in way we reason to our judgments in these different domains, we have a poor understanding of the content of the moral truths whereas we have had a much better understanding of the content of perceptual truths and mathematical truths—that is why it is no surprise that we do not

yet have a good of moral reliability even though we have had good explanations of perceptual and mathematical reliability for a long time. If the objection is construed this way, it should be pointed out that the moral realists targeted by the *Explanatory Challenge* are epistemic optimists: they think we are justified in holding many of our moral judgments, that we have a good grip on a wide range of plausible moral truths (or approximate truths), not just a few truths. So the targeted moral realists are not well positioned to invoke this objection in defense of the *Explanation from Insufficient Effort*. If they do so, they do so at the cost of adopting an epistemically impoverished and implausible form of moral realism.

5.4.3 Against Cognitive Closure and Evolutionary Handicap

Now consider the *Explanation from Cognitive Closure*, i.e. whether we are cognitively closed to a plausible explanation of moral reliability. If invoked to explain the realist's failure to plausibly explain moral reliability, the bare possibility of cognitive closure appears *ad hoc*. If we had good reason to think a plausible explanation is *in principle* epistemically inaccessible to us, then we should of course stop pointlessly striving for an explanation. But the fact that moral philosophers have been busily searching for an explanation suggests that they see no good reason to accept cognitive closure. I also see no good reason to accept it. After all, we have plausible explanations

of reliability in other domains (e.g. perceptual, scientific, mathematical, and perhaps various philosophical domains). So we need reason to presume that something special about the moral domain makes it impossible for creatures like us to explain reliability in this domain. What specific features of our cognitive faculties or explanatory capacities put moral reliability beyond their possible reach? Without a convincing story, the bare claim that we have faculties that could not possibly explain moral reliability is highly doubtful.

Colin McGinn (1993, Ch. 6), however, sketches a story for us. According to McGinn, it is the abstract (non-spatiotemporal, causally inert) character of the moral facts that makes it impossible for spatio-temporally situated creatures like ourselves to explain how we are reliably detecting them. The basic idea is that though our evolved causal-spatial mode of awareness allows us to explain empirical knowledge by appeal to causal and spatio-temporal naturalistic relations, no such naturalistic relations obtain in cases of *a priori* knowledge where the subject matter is abstract. Hence, we are closed off by our causal-spatio mode of awareness from a plausible explanation of moral reliability on a robust non-naturalist understanding of the moral facts.⁶

⁶ McGinn (1993), see Ch. 6, esp. pp. 99-102; also see pp. 18-20, 101-102, 128-131, and McGinn (1994), pp. 146-151.

Two points suggest McGinn's story does not help the moral realist out. First, McGinn's explanation would at best only explain why we have no plausible explanation of moral reliability on Platonist or non-naturalists view of the moral facts where the moral facts are characterized as non-spatiotemporal and causally inert. So McGinn's explanation, even if convincing, could only defend a robust non-naturalist version of moral realism from the *Explanatory Challenge*.

Second, McGinn's explanation is unconvincing. For one thing, many philosophers think we are not cognitively closed to a plausible explanation of *mathematical* reliability—indeed many think we indeed possess a plausible outline of an evolutionary explanation of mathematical reliability despite the abstract (nonspatio-temporal, causally inert) nature of the mathematical facts. The basic idea is that we evolved a disposition to get the basic mathematical facts right because tending not to get them right would be disastrous to our survival.⁷ For example, we evolved to believe that $1+1=2$ because our ancestors and their cognitive equipment would not have survived if they believed that $1+1\neq 2$. So then, since it possible for us to come up with a plausible explanation of mathematical reliability despite the abstract nature of the mathematical facts, it is *at least possible* that we could come up with a plausible

⁷ Sosa (2002), Joyce (2006, p. 182), and De Cruz (2007).

explanation of moral reliability on an abstract understanding of the moral facts. Hence, the claim of cognitive closure—that is, the claim that it is *impossible* for us to come up with a plausible explanation of moral reliability—appears dubious.

Now consider the *Explanation from Evolutionary Handicap*. Explanation (5) sketches a story for thinking not that it is *impossible* but that it is *unlikely* that we would be able to explain moral reliability at this point. The basic idea is that our brains evolved to help us survive and reproduce and so we should not expect them, at least at the present moment, to enable us to explain something like moral reliability. Rather, it is likely that we are cognitively handicapped by evolution with regard to this explanatory matter (though perhaps not cognitively closed). Though weaker and more plausible than the invocation of cognitive closure, the invocation of this *local* evolutionary handicap on our part appears quite difficult to flesh out and substantiate, which could explain why nobody has done it yet.⁸

Our evolutionary heritage undoubtedly imposes cognitive limitations on us, but *in which domains* are we handicapped by evolution such that explanations in these domains are likely beyond our present reach? We (or the relevant experts) currently possess astonishing capacities to do higher-level mathematics, quantum physics, and

⁸ For discussion of how our evolved brains limit our explanatory capacities, see Dawkins (2006, pp. 367-74). For related discussion, see McGinn (1993, 1994).

explain much of the world around us. For example, we have come up with plausible explanations of our perceptual, scientific, and mathematical reliability. And we have made substantial progress regarding many other ordinary, scientific, and philosophical explanatory questions as well. For example, though our evolved brains suggest to us that tables are solid objects through and through, science has discovered that tables are almost entirely composed of empty space that is dotted by particles located at surprising distances from each other. On the philosophical side, many think we have discovered that the nonexistence of the God of traditional theism is part of the best explanation of the kinds and extent of evil in the world, that the falsity of a belief explains why it does not count as knowledge, and so on. The capacities to answer such explanatory questions are surely not essential to survival, but rather are byproducts or extensions of our other selected cognitive capacities, like our domain-general capacity for reasoning and our inductive reasoning capacities. Given how much we can explain in other domains and that our explanatory capacities in these domains are not essential to survival, it is at least unclear what could justify the suspiciously selective expectation that our cognitive capacities would not allow us to explain moral reliability, especially when we have a good handle on plausible moral belief-forming processes and plausible moral truths, have searched long and hard for the explanation, and so on.

Absent a principled account of what explanatory questions we cannot expect our evolved brains to answer—and thus why we should not expect ourselves to be able to explain the local matter of moral reliability, even though we can plausibly explain much else—it appears only a *generalized* sort of explanatory pessimism could ground *local* explanatory pessimism with regard to moral reliability. But then such pessimism, given its generalized character, would most likely press us to undesirable skeptical consequences elsewhere—for instance, to explanatory skepticism with regard to large portions of the philosophical domain. So the explanation by appeal to a generalized evolutionary handicap on our part does not appear promising, at least if one is modestly optimistic about our explanatory capacities in the philosophical domain.

An additional point should be made: even if certain moral realists happen to be deeply pessimistic about our explanatory capacities in the philosophical domain, they nonetheless invariably rely on their explanatory capacities when they champion the explanatory power of moral realism—that is, the explanatory power of response-independent moral facts. But whatever generalized pessimism grounds local pessimism about our capacity to explain moral reliability would seem to also ground pessimism about our capacity to explain (say) the intuitive categorical normativity of the moral facts by invoking their realist nature. The idea is that if moral realists are pessimistic about how much our evolved brains can be expected to explain in the philosophical

domain, this would undercut the force of their arguments for moral realism as the best explanation of various purported phenomena (e.g. like categorical normativity). So though moral realists *could* invoke a generalized pessimism about our explanatory capacities, in so doing they would undercut the force of the arguments that primarily motivated moral realism in the first place. This is an unstable strategy.

5.4.4 Against Bruteness and Chance

Now consider *Bruteness Explanation*. The process reliability of our moral judgments is not plausibly a metaphysically brute, unexplainable fact. Plausible candidates for brute facts consist of such things as the basic constituents of reality or the fundamental laws of nature, or perhaps even the existence of basic necessary moral facts, but the process reliability of our moral beliefs does not resemble such candidates in a way that might give us reason to think they are similarly brute. For example, though basic moral facts or standards might be necessary and perhaps brute, the process reliability of our moral judgments is undoubtedly a contingent fact: the processes behind our moral judgments could have been unreliable. Claims of bruteness beyond necessary facts or the basic constituents or laws of the universe appear overreaching, and the claim that moral reliability is brute is no exception.

As for *Chance Explanation*, it is hard to believe that moral reliability is a product of chance. The moral reliability of the processes behind our moral judgments is a striking alleged fact that, if it obtains, “calls out” for an explanation, just as much as our perceptual reliability does or our mathematical reliability does. Though this is not the place to develop an account of strikingness or of the conditions under which an alleged phenomenon merits an explanation, we can nonetheless appreciate the strikingness of moral reliability⁹—on a standard probabilistic understanding of reliability—by observing that it amounts not just to an actual world correlation but rather to a *counterfactually stable* correlation between our moral judgments and moral facts *that are fixed independently of our judgments and attitudes*. Though an actual world correlation between a (non-gerrymandered) domain of beliefs and its corresponding realist facts may sometimes plausibly be the result of chance—some individuals or groups might just be lucky believers with respect to some response-independent domain—it is difficult to believe that a *counterfactually stable* correlation between a domain of beliefs and its corresponding realist facts is the result of chance. I cannot think of a plausible

⁹ Relevant analogy: we can know that an alleged phenomenon merits an explanation even if we have no general account of the necessary and sufficient conditions under which a phenomenon merits an explanation, just as we can know that we know something even if we have no account of the necessary and sufficient conditions for knowledge.

example of such a correlation being due to chance.¹⁰ Just like the strong counterfactually stable correlations between smoking and lung cancer or between high tides and the moon's proximity to the earth, the correlation between our moral judgments and realist moral facts calls out for an explanation. The correlation could possibly be an accident, but from an epistemic point of view it is extremely unlikely. All this is to say: the chance hypothesis strains credulity.

5.4.5 *The Best Explanation: Eliminativism*

The argument so far suggests that the candidate explanations suggested above are not good explanations of *e*. Rather, the best explanation of *e*, of why realists have no good explanation of moral reliability, seems to be the *Eliminativist Explanation*: there is no moral reliability. Not only does it seem to be best, reasons to think so are provided its scores on the following general (non-contextual) criteria of explanatory merit:

Ontological simplicity: ceteris paribus, explanations that posit fewer entities and fewer kinds of entities are better.

¹⁰ The only remote candidate might be something like chicken-sexer reliability. But even in this case there is a plausible explanation for how chicken-sexers can reliably determine the sex of chicken hatchlings, even if they themselves have no reflective access to it. See Horsey (2002) for such an explanation.

Explanatory simplicity: ceteris paribus, explanations that are elegant, raise fewer difficult questions, and operate with fewer primitive notions are better.

Non-adhocness: ceteris paribus, explanations with fewer *ad hoc* components are better.

Explanatory scope: ceteris paribus, explanations that account for a wider range of phenomena are better.

Fruitfulness: ceteris paribus, explanations that yield more novel predictions are better.

Testability: ceteris paribus, explanations that have more testable consequences are better.

In what follows, it is argued that the *Eliminativist Explanation* scores well on such criteria of explanatory merit. An important qualification must be made the outset: the *Eliminativist Explanation* scores well *with respect to* the particular *explanandum* of the moral realist having no plausible explanation of moral reliability. The holistic and infeasible question of how the *Eliminativist Explanation* fares on explanatory criteria *with respect to all explananda* (e.g. all the *explananda* of moral discourse and practice) will not be addressed here.

The *Eliminativist Explanation* posits no new entities or kinds of entities—for example, it does not posit a domain of moral facts, nor new kinds of special belief-forming faculties or processes—hence its ontological simplicity. It is rather straightforward, elegant, and explanatorily simple. It is easily testable and makes the novel prediction that all extant and future explanations of moral reliability will prove implausible. It is not *ad hoc* in the sense that the fact that it is invoked to account for the *explanandum* is not the *only* reason to accept the claim that there is no reliability¹¹: other reasons to doubt reliability are provided by formidable philosophical arguments.¹² Moreover, it is authentically puzzling why we should rely on or trust the processes of biological and cultural evolution to lead us to moral truth when that truth is fixed independently of our responses. So nothing appears *ad hoc* about the *Eliminativist Explanation*.

The *Eliminativist Explanation* also explains a wide range of phenomena. For one thing, given that realists have no plausible explanation of moral reliability, it accounts for the failure of all the various past and present realist attempts to explain moral reliability (see the previous chapter for the case that realists have no explanation of moral reliability). It also explains why we find so much disagreement in the area—that

¹¹ For this conception of ad-hocness, see Vogel (1990, p. 659).

¹² E.g. Sinnott-Armstrong (2008).

is, nothing resembling convergence on an explanation, even a broad sort of explanation. It also explains why throughout the history of moral philosophy we see the same standard explanations (e.g. Rationalist explanations, see the previous chapter of this dissertation) come in fashion, then go out of fashion, then come back in—as opposed to seeing the progressive elimination of explanations that cannot be made to work. All this is to say: the explanatory question seems intractable and the *Eliminativist Explanation* neatly explains why.

The considerations above suggest that the *Eliminativist Explanation* is not simply a better explanation than its rivals, it is a positively *good* explanation that scores well on accepted standards of explanatory merit. But are there common explanatory criteria on which the *Eliminativist Explanation* does not do so well? Two candidates suggest themselves:

Coherence with Background Knowledge: ceteris paribus, explanations that cohere with our background knowledge are better.

Conservatism: ceteris paribus, explanations that better preserve common and entrenched beliefs are better.

The objection that the *Eliminativist Explanation* does not cohere with our background knowledge is a serious charge. For even if the explanation scores high on other measures of explanatory merit, we could still have sufficient reason to reject it if it would imply that we need to revise well-established background theories and beliefs. To illustrate the principle: even if the existence of a non-physical soul would best explain certain phenomena like near-death and out-of-body experiences, this plausibly still would not provide us with sufficient reason to accept that we have non-physical souls, for positing such entities would require us to revise much of our well-established background theories in physics and biology.¹³ But the principle does not apply in our case. The claim that the *Eliminativist Explanation* conflicts with established background knowledge either begs the question at issue—that is, assumes moral reliability is part of our background knowledge—or needs to be supported by good arguments that run from premises provided by our background knowledge to the conclusion of moral reliability. If the needed premises are provided by non-moral background theories, then such arguments require crossing the difficult is-ought gap. If the needed premises are moral, it is highly debatable whether there are good arguments from moral premises to

¹³ See Carruthers (2004), pp. 144-146.

the conclusion of moral reliability.¹⁴ Hence, the claim that moral reliability is part of our background knowledge is suspect. So the objection from incoherence with our background knowledge does not get off the ground.

However, the criterion of *Conservatism*—which makes no claim concerning the *epistemic status* of common and entrenched beliefs—appears to knock against the *Eliminativist Explanation*, since moral reliability appears to be widely accepted. Two points suggest, however, that the knock is at best a slight bump. First, though belief in basic moral truths is “common and entrenched,” it is at least unclear that the belief in process reliability on realist views of the moral facts is a common and entrenched belief. Recent empirical studies suggest that relativist (i.e. response-dependent) views of the moral facts are quite common.¹⁵ Second, even if the vast majority of people accept an objectivist or realist metaethics and buy process moral reliability, this does not impugn the explanatory merit of the *Eliminativist Explanation*. For there are always trade-offs among explanatory criteria and *Conservatism* with respect to the *philosophical domain* (setting aside core principles of logic/reasoning) is not among the weightier criteria. Conservation of common and deeply entrenched *scientific* background theories and

¹⁴ The few extant realist attempts to vindicate moral reliability (see e.g. Shafer-Landau 2001) are, in my view, far from convincing.

¹⁵ Goodwin and Darley (2008), Sarkissian et. al. (2011).

principles of logic/reasoning is arguably a weighty criterion. But moral reliability does not lie within these domains. Given that the *Eliminativist Explanation* does quite well on other standard and weightier criteria like explanatory scope, it is still a positively good explanation even if it conflicts with the alleged common philosophical belief in moral reliability (on moral realism). So though explanatory assessment is never an exact science, we have good reason for thinking the *Eliminativist Explanation* is, given the comparatively poor quality of the alternative explanations, the best explanation of the realist failure to explain moral reliability.

5.4.6 *Inference to the Best Explanation: What Follows?*

Even if the *Eliminativist Explanation* is best, however, of what epistemic significance is this? Does it follow that the explanation is probably true? That is, in terms of the *Explanatory Challenge*, how can we reasonably draw a skeptical conclusion from premise (9)? Discussion of the epistemic significance of IBE is in order.

There are various versions of the IBE rule or principle. The standard version runs as follows:

Standard IBE Principle: Given *explanandum* e and available alternative explanations $H_1 \dots H_n$, if hypothesis H_i would (if true) best explain e , then H_i is probably true.

This sort of inference is widely used and unquestioned in most areas of our personal and intellectual lives. Natural philosophical questions, however, surround its epistemic status. It obviously is not a logically sound form of inference, but is it epistemically reliable? If so, to what degree? That is, when its premises are true, is its conclusion likely to be true? If so, how likely? In which domains? The reliability question is important because on it hinges whether and in what domains we should believe our best explanations (e.g. *Eliminativist Explanation*) to be true explanations.

Though classic questions surround the alleged truth-conduciveness of standard explanatory virtues—alternatively put, whether explanatory “loveliness” is positively correlated with probability¹⁶—the most important and prominent reason for doubting the reliability of IBE is given by the “best of a bad lot” objection.¹⁷ This objection begins by pointing out that IBE is only reliable, or likely to lead to an approximately true explanation, on the supposition that the approximately true explanation is likely to

¹⁶ See Lipton (2004) for discussion.

¹⁷ Van Frassen (1989), pp.142-143.

reside among the explanations up for consideration. For if the true explanation is often not among the set considered, then the best explanation of the set considered often may well be the “best of a bad lot,” i.e. the best among highly improbable explanations.

The objection then observes that when we employ IBE we seldom if ever consider *all* potential explanations of the phenomenon in question. Rather, given our cognitive and temporal limitations, we generate a shortlist of candidates and then proceed to compare their merit according to standard explanatory criteria. The question that bothers is this: given that we cannot feasibly consider all potential explanations, why think the true explanation is likely to be among the batch we consider? It is difficult to see what would vindicate this assumption, especially in light of our cognitive limitations and the historical contingencies that undoubtedly affect which explanations we take seriously. The assumption that the true explanation resides among the ones we come up with appears to be a piece of arbitrary epistemic self-privileging. But the assumption is essential to vindicate the reliability of IBE. The objection concludes: since the assumption is implausible, IBE is unreliable.

The objection from bad lots suggests that for the IBE rule to be a reliable inference rule, it should be amended this way:

Revised IBE Principle: Given *explanandum* e and available explanations $H_1 \dots H_n$, if hypothesis H_i would (if true) best explain e and is not the best of a bad lot, then H_i is probably true.

My argument takes the objection from bad lots seriously and thus operates with an IBE rule incorporating the “no bad lots” condition. It must be argued, accordingly, that the best explanation of the realist’s failure to plausibly explain reliability—namely, the *Eliminativist Explanation*—is not the best of a bad lot. Given that the objection from bad lots may impugn the reliability of IBE in some domains and not others, we need not argue that the scientific domain, philosophical domain, or any other general domain is typically free of bad lots. Rather, we need to argue that *with regard to this particular explanatory question* the *Eliminativist Explanation* is not the best of a bad lot. Our task then is easier than the task of those who would ambitiously defend the reliability of IBE in science, philosophy, or other generalized domains.

The claim that the *Eliminativist Explanation* is not the best of a bad lot—because it is not only the best explanation but a positively good explanation—gains its main support from the earlier argument for the intrinsic (rather than merely comparative) explanatory merit of this explanation. If the *Eliminativist Explanation* scores well on standard criteria of explanatory merit and does not score too poorly on other standard

criteria, this gives reason for thinking it is not merely the comparatively best, but a positively good explanation. Granted, there is no algorithm for determining the threshold of explanatory merit that demarcates the probable from the improbable—such thresholds elude—but the earlier motivation of the intrinsic explanatory merit of the *Eliminativist Explanation* pushes it towards the “probable” end of the spectrum and thus significantly mitigates, if not entirely deflates, the objection from bad lots.

So then, according to the *Revised IBE Principle*, the *Eliminativist Explanation*, as the best explanation of a good lot, is probably true everything else equal. But a further question remains: even if the best explanation is probable, *how* probable is it?

Alternatively put: even if IBE is reliable in a given domain—in the sense that it at least implies the best explanation is more likely to be true than not—*how* reliable is it? This question is important since the mere fact that an explanation is “probably true” (i.e. more than 0.5 likely to be true) does not reasonably imply that we should accept or believe the explanation to be true. That is to say: it is not clear that the IBE argument, all by itself, can achieve the degree of probability requisite for justified belief in the truth of the best explanation.

What is needed for my IBE argument to imply a skeptical conclusion, however, is not that we are justified in *believing* the best explanation—the *Eliminativist Explanation*—but rather that we should *suspend judgment* regarding its truth. For if we should

suspend judgment regarding *the Eliminativist Explanation*—that is, regarding the claim that there is no moral reliability—this means we should suspend judgment regarding moral reliability. Justified suspension of judgment regarding moral reliability in turn would constitute a significant defeater of whatever initial justification our moral beliefs might have enjoyed. This would be an important skeptical upshot.

But can an inference to the best explanation (free of bad lots) get us to justified suspension of judgment regarding moral reliability? Yes. For since inference to the best explanation (free of bad lots) implies the best explanation is probably true—that is, at least more likely to be true than not—it *at least* warrants us in suspending judgment about the truth of the best explanation, if it does not warrant us in believing its truth outright. Alternatively put: inference to the best explanation (free of bad lots) implies the disjunctive claim that we should either suspend judgment about the best explanation or accept it (we should not deny it). With this last iteration we have arrived at the sophisticated IBE principle that actually underlies the IBE argument of the *Explanatory Challenge*. The principle is located at premise (11) and runs thus:

(11) *IBE principle*: Given *explanandum* e and available explanations $H_1 \dots H_n$, if hypothesis H_i would (if true) best explain e and is not the best of a bad lot, then we should either suspend judgment regarding H_i or accept H_i .

The IBE argument of the *Explanatory Challenge* thus runs as follows: since the *Eliminativist Explanation* (i.e. the fact that there is no reliability) would best explain the moral realist's failure to explain moral reliability and is not best of a bad lot, then it follows that we should either suspend judgment regarding the truth of the *Eliminativist Explanation* or accept it.

5.4.7 Reaching the Skeptical Conclusion

The IBE argument from (1) to (12) in the *Explanatory Challenge* appears in good standing. We should either suspend judgment regarding moral reliability or accept that there is no moral reliability. But this could be too quick. A "companions in guilt" objection is suggested by the worry that we have no plausible explanations of mathematical reliability on realist (e.g. Platonist) understandings of these facts. If we have no such explanation, then my challenge appears to imply a contentious reliability skepticism on such realist views by implying that the best explanation of our failure to explain mathematical reliability is that there is no mathematical reliability.

But my challenge implies no such thing. This is so for two reasons. First, in the case of our mathematical beliefs, as pointed out earlier, we arguably do possess in broad outline a plausible evolutionary explanation of our mathematical reliability. Second, a

rear guard point: even if we do not possess a plausible evolutionary explanation of mathematical reliability—for example, if the best explanation we have is so sketchy and incomplete that it does poorly on standard criteria of explanatory merit—then it is highly doubtful that premise (2) is satisfied with regard to our mathematical beliefs. Though we have plausible explanations, at least in broad outline, of the origins of our moral beliefs, we have a comparatively much weaker grip on the sorts of processes behind our mathematical beliefs.¹⁸ This implies that a live explanation of our failure to explain mathematical reliability—if indeed we have failed—is that we have little to no idea where these judgments come from. That is, it is opaque in the mathematical domain whether an eliminativist explanation is better than an explanation from causal ignorance. In the moral case, however, it appears the *Explanation from Causal Ignorance* is not the best explanation of the realist's failure to explain moral reliability, for empirical moral psychology is comparatively more developed than mathematical psychology. (For more on this point, see the next section for an extended response to an objection to premise (2), namely the objection that empirical moral psychology is not in good epistemic standing and thus the best explanation of our failure to explain moral reliability (on moral realism) is the *Explanation from Causal Ignorance*).

¹⁸ E.g. see De Cruz (2007) on the origins of our mathematical beliefs.

Setting aside companions in guilt objections, the inference to the skeptical conclusion that we should either suspend judgment regarding moral reliability or accept that there is no moral reliability may also be too quick for another reason. We cannot yet accept the conclusion until we give thought to whether there are *overriding* reasons that weigh in favor of accepting moral reliability. For if there are weighty reasons to accept moral reliability, then the fact that the *Eliminativist Explanation* is best does not by itself defeat moral reliability. Consider analogies. The fact that our recent ancestors did not have a plausible explanation of the correlation between smoking and lung cancer or between high tides and the moon's proximity to the earth did not imply there was no such correlation or that they should suspend judgment regarding such a correlation, precisely because they had a lot of independent evidence that there in fact was such a correlation. Similarly, if we have good independent evidence for moral reliability, the fact that we cannot yet plausibly explain it would not imply there is no such reliability or that we should suspend judgment regarding such reliability. So the question is whether moral realists have such overriding reasons. Premise (13) presumes that they do not.

Given the present state of philosophical debate, it is a blurry if not entirely opaque matter what these overriding reasons could be. Though this is not the place to defend the claim that moral realists have no such good reasons by critiquing extant

realist arguments for moral reliability, it is important to observe that though moral philosophers have given sustained attention to coherentist, intuitionist, and explanationist¹⁹ ways of vindicating the justification of our moral judgments, extremely little attention has been paid to reliabilist support.²⁰ The few serious attempts to vindicate moral reliability on moral realism (e.g. Shafer-Landau 2001) are, in my view, far from convincing.

So then, on the presumption that there is not a successful vindication of moral reliability available that would override the reliability undermining secured by the IBE argument of the *Explanatory Challenge*, it follows in (15) that realists should either suspend judgment regarding moral reliability or accept that there is no moral reliability (premise (14) expresses the inference principle). If this is so, then we can reasonably infer in (17) that moral realists have good reason to think they are not justified in holding their moral beliefs (premise (16) expresses the inference principle). For if we discover that we should suspend judgment regarding or deny the reliability of a belief forming process, that provides a significant defeater of the justification of the process's output

¹⁹ "Explanationist" routes to supporting our moral judgments maintain that our moral judgments are justified because the moral facts that make them correct play a role in their best explanation. For a nice review of the explanationist literature, see Majors (2007).

²⁰ Shafer-Landau (2001) points this out and goes on to offer interesting but to my mind unconvincing arguments on behalf of the reliabilist justification of some of our moral judgments.

beliefs. Perhaps we cannot yet make an all-things-considered judgment until we rule out overriding reasons the moral realist might have for thinking her moral beliefs are true. That is, even though the moral realist may be justified in suspending judgment regarding or denying the reliability of the sources of her moral beliefs, nonetheless she might step back and find a way to independently test these beliefs to determine their truth. For example, the moral realist might argue that even though our moral judgments are unreliably produced, at least some of them are justifiably held on the grounds that their corresponding moral facts play a role in the best explanation of our moral judgments. That is, moral realists might still avoid moral skepticism by a successful explanationist argument of the sort championed by Nicholas Sturgeon (1984) and others.

The *Explanatory Challenge* concludes with an appropriately modest but still significant *ceteris paribus* skeptical claim in (16) that moral realists have *good reason* to deny the *Epistemic Claim* that we are justified in holding at least some moral judgments. It identifies a big mark against the justification of our moral beliefs (on moral realism). This is a big mark, since if all of our moral beliefs are produced by processes the reliability of which is in doubt, it is doubtful that we can step back and verify them. *Prima facie* justification conferred by moral phenomenology or moral experience or “self-evidence” upon our moral beliefs is reasonably defeated when we find out that we

should suspend judgment regarding the reliability of their sources.²¹ So then, the *ceteris paribus* skeptical claim seriously threatens to lead to moral justification-skepticism full stop with regard to all our moral beliefs (on moral realism). If the *Explanatory Challenge* can successfully press us this far, it is the strongest version of the explanatory challenge developed thus far.

5.5 Objection: What Do Empirical Moral Psychologists Know?

5.5.1 The Objection

An anticipated objection to premise (2) of the *Explanatory Challenge* is driven by pessimism regarding the epistemic standing of empirical moral psychology (i.e. the empirical work done in a variety of subdisciplines recognized within psychology that deal with matters relating to morality). The objection runs: the psychological sciences give us this vague sense that it is partly nature and partly nurture but not much more than that—psychologists at best have a grip on some very partial and nebulous causes of

²¹ What about coherentist sources of justification? One could argue that it is doubtful that coherence among moral judgments the reliability of which is in doubt confers much (if any) justification upon them (garbage-in-garbage-out). Moral coherentists have a response available, namely that the coherence of unreliably formed moral judgments with each other *and with reliably formed non-moral judgments*—that is, coherence in *wide* reflective equilibrium—could confer justification upon our moral judgments. In future work I aim to engage this response. In the present argument, however, all that I am arguing is that we have good initial evidence—which may or may not ultimately be overridden by coherentist sources of evidence—for thinking that we are not justified in holding our moral judgments.

our moral judgments. Psychologists have different views about what the plausible operative partial causes are, but regardless, the objection goes, they have only a few very small and blurry pieces of the causal puzzle. If this is so, the *Explanatory Challenge* may be defused, for the best explanation of the realist failure to plausibly explain moral reliability could then very well be that nobody has a grip on plausible causes behind our moral judgments, that is, the *Explanation from Causal Ignorance*.

5.5.2 *Assessing the Objection*

Short of an ambitious assessment of the scientific and epistemic status of mainstream research programs in empirical psychology, the best response to this sort of objection by a naturalistically-inclined philosopher, in my view, is to defer to empirical moral psychologists. What do they think they know? After all, who are philosophers to tell empirical moral psychologists that they have no idea what the causes of our moral judgments might plausibly be? The scientists are better positioned to address the matter. Though empirical moral psychologists do not speak unanimously as one group, many of them, the vast majority one suspects, are rather optimistic about the explanatory power of mainstream research programs in their discipline and think we have a good handle on the broad outlines of plausible explanations of our moral judgments. For instance, moral psychologists not infrequently suggest that we have a

good handle on the affect-laden intuitive processes and reasoning-based processes behind our moral judgments,²² many evolutionary theorists think they have a good handle on plausible evolutionary sources (e.g. innate biases, processes of cultural evolution)²³, and the same goes for many social psychologists with regard to an array of socialization processes (e.g. modeling, conformist pressures). Put the outlined causal components together and we have a fairly plausible explanation, at least in broad outline.

The point is perfectly compatible with the recognition that some research programs in psychology—especially those outside the academic mainstream—are dubious (say, Freudian psychiatry) and with the recognition that the details of plausible explanations need much filling in and that the mechanisms are often contested and not well-understood. The point is also perfectly compatible with the recognition that mainstream empirical psychology, as a comparative matter, has not experienced as much progress and is not as epistemically credible as “hard” sciences like physics and chemistry. Finally, the point is also compatible with the suggestion that many

²² See e.g. Haidt’s (2007) article in *Science*—entitled “The New Synthesis in Moral Psychology.” He discusses the convergence of moral, social, and evolutionary psychology on several components of the best explanation of our moral judgments. Also see Nichols (2004).

²³ See e.g. Levy (2004), Boyd and Richerson (2005), Joyce (2006), and James (2011).

psychologists, perhaps the vast majority, are probably too overconfident in the assessment of the epistemic status of their enterprise. The crucial point is that as long as mainstream empirical psychology is sufficiently credible such that we have a plausible (albeit rough and gappy) explanatory picture of the origins of our moral judgments, then the central IBE inference of the *Explanatory Challenge* is plausible. Moreover, the inference gets stronger the clearer and more extensive a picture we get regarding the psychological sources of our moral judgments.

If the moral realist digs in her heels because she is a scientific anti-realist or a selective anti-realist who denies the epistemic credibility of mainstream empirical moral psychology, then, in my view, she is standing on questionable ground. Of course, this is not the place to defend scientific realism or the scientific and epistemic status of mainstream empirical moral psychology—an ambitious project—so we may consider my challenge to be conditional: *if* mainstream empirical moral psychology provides us with plausible explanations, at least in broad outline, of the origins of our moral judgments, then skeptical implications reasonably follow along the lines suggested by the *Explanatory Challenge*. And the more plausible and complete these psychological explanations become, the more reasonable it is to suppose these skeptical implications follow.

An objector could grant that we have a broad and rough plausible picture of the origins of our moral judgments but nonetheless press that this picture is *too rough*—that we should not expect to have an explanation of moral reliability without more complete and precise models of the relevant causes. But it is at least unclear why we should *expect* to have a good explanation of moral reliability once we obtain more exhaustive and precise models.

Compare the current state of empirical moral psychology with the progressive grip that we (including our post-Darwin intellectual forebears) have obtained on the causes of our *perceptual* beliefs and *mathematical* beliefs. We did not need a finely detailed picture that psychologists unanimously agreed about in order to have the broad outline of a good explanation of our perceptual reliability: with an evolutionary understanding of our perceptual psychology in place, a good explanation naturally suggested itself very early on after Darwin. To my knowledge, today we *still* have neither a finely detailed picture nor any sort of consensus on the psychological mechanisms generating our perceptual beliefs. Yet we nonetheless possess (as our post-Darwin forebears possessed) a broad plausible explanation of perceptual reliability that would defuse the *Explanatory Challenge* if it were aimed at perceptual reliability rather than moral reliability. Similarly, we have no finely detailed picture of the origins of our mathematical beliefs, but we nonetheless have a plausible broad explanation of our

mathematical reliability that would reasonably defuse the *Explanatory Challenge* if it were aimed at mathematical reliability (on realist understandings of the mathematical facts).²⁴

The above considerations naturally raise the question: if we did not need finely detailed and exhaustive causal accounts of our perceptual beliefs and mathematical beliefs in hand to come up with a good explanation of our perceptual reliability and mathematical reliability, why suppose that once moral realists have such accounts of the origins of our moral beliefs they will likely come up with a good explanation of moral reliability? No good answer appears ready to hand. Given that we did not need better causal accounts of the origins of our perceptual and mathematical beliefs to allow us to explain perceptual reliability and mathematical reliability, it appears that it is the objector's burden to show why with better causal accounts of our moral beliefs we will likely come up with a good explanation of moral reliability.

Of course it is *possible* that with more complete and precise causal models in hand realists will obtain the desired explanation, just as it is possible that they would obtain a plausible explanation from the emergence of a realist Newton for moral

²⁴ The explanation of mathematical reliability in mind is that we are evolutionarily disposed to form true basic mathematical beliefs because tending to form false mathematical beliefs would be disastrous for our survival. For example, we evolved to believe that $1+1=2$ because our ancestors and their cognitive equipment would not have survived if they had instead believed that $1+1=0$. On the evolutionary explanation of mathematical reliability in mind, see Sosa (2002), Joyce (2006, p. 182), and De Cruz (2007).

epistemology. But the relevant question is whether it is likely to happen. That is, is it *likely* that by doing more and better psychology—say, by filling in one of the gaps in our explanatory picture—there will emerge a plausible explanation of moral reliability (on moral realism)? Or is it *likely* that with the emergence of a really smart moral realist (a “Newton”), she or he will finally come up with a plausible explanation for us? I see no reason for accepting optimistic probability estimates on the matter. Rather, given that very smart philosophers have tried hard for a long time to come up with such a realist explanation of moral reliability, given that they have had a fairly good handle on the origins of our moral judgments, given that they have had a good handle on plausible moral truths, and given the implausibility of invocations of cognitive closure, evolutionary handicap, metaphysical bruteness, and chance, it is reasonable to suppose that if realists have not been able to connect the dots so far, the dots are in all probability not connected. That is, though we of course cannot disprove the possibility that a plausible explanation of moral reliability will emerge, we should nonetheless go where our current evidence leads. And our current understanding of the subject matter suggests that unless there is an unexpected scientific or intellectual revolution as dramatic as Darwin’s that finally allows us to solve the explanatory puzzle, it is unlikely that a plausible explanation for moral reliability will emerge.

5.5.3 A Local Version of the Explanatory Challenge

Let me develop one further response to the objection from psychological pessimism, one that appears quite philosophically interesting. *Even if* most psychologists are pessimistic about their discipline's explanatory power, and even if our best psychological sciences only give us a few very partial and very rough explanations of our moral beliefs (that we cannot put together to get a sufficiently comprehensive picture), nonetheless an interesting *local version* of the *Explanatory Challenge* still possesses significant skeptical bite.

The local version of the challenge runs as follows. Observe that there are multiple contributory causes or particular influences that play a causal role in the processes leading to our moral judgments. For these influences—taken singly or jointly—it is sensible to ask whether they are reliability-conferring (truth-tracking) or reliability-detracting (distorting). For the sake of argument, grant that the moral realist has no plausible explanation of the reliability of particular contributory causes X, Y, and Z, which play a causal role in the processes behind all, many, or some of our moral judgments. For causes X, Y, and Z we might plug in standard causes such as certain evolutionary biases, emotions, processes of cultural evolution, and so on. What could explain this local explanatory failure on the part of moral realists? In sync with our

earlier discussion of the generalized version of the challenge, consider the following natural candidate explanations:

*Explanation from Causal Ignorance**: Philosophical experts—i.e. smart philosophers who have tried hard to explain moral reliability—are not aware of the best causal accounts of X, Y, and Z as causes of our moral beliefs.

*Explanation from Moral Ignorance**: Philosophical experts are not aware of plausible moral truths.

*Explanation from Insufficient Effort**: Philosophical experts have not tried long and hard enough to explain moral reliability on the best empirical accounts of X, Y, and Z as causes of our moral beliefs and on plausible understandings of moral truth.

*Explanation from Cognitive Closure**: Philosophical experts are “cognitively closed” to plausible explanations of the moral reliability of X, Y, and Z-- i.e. such explanations are in principle inaccessible to creatures like us.

*Explanation from Evolutionary Handicap**: Philosophical experts are cognitively handicapped by evolution in their search for plausible explanations of the moral reliability of X, Y, and Z.

*Bruteness Explanation**: The moral reliability of X, Y, and Z is a brute fact, hence unexplainable.

*Chance Explanation**: The moral reliability of X, Y, and Z is due to chance, hence unexplainable.

*Eliminativist Explanation**: X, Y, and Z are not morally reliable.

The local version of the *Explanatory Challenge* can then be spelled out thus:

Explanatory Challenge (Local Version)

Let *e* refer to the fact that the moral realist has no plausible explanation of the moral reliability of contributory causes X, Y, and Z, which play a causal role in the processes leading to our moral judgments.

- (1) *e* obtains. I.e., the moral realist has no plausible explanation of the moral reliability of X, Y, and Z.
- (2) *Explanation from Causal Ignorance** is a poor explanation of *e*.
- (3) *Explanation from Moral Ignorance** is a poor explanation of *e*.
- (4) *Explanation from Insufficient Effort** is a poor explanation of *e*.
- (5) *Explanation from Cognitive Closure** is a poor explanation of *e*.
- (6) *Explanation from Evolutionary Handicap** is a poor explanation of *e*.
- (7) *Bruteness Explanation** is a poor explanation of *e*.
- (8) *Chance Explanation** is a poor explanation of *e*.
- (9) *Eliminativist Explanation** is the best explanation of *e*. I.e., the fact that X, Y, and Z are not morally reliable would (if true) best explain why the moral realist has no plausible explanation of their reliability.
- (10) The set of explanations considered is not a bad (highly improbable) lot of explanations.
- (11) *IBE principle*: Given *explanandum e* and available explanations $H_1...H_n$, if hypothesis H_i would (if true) best explain *e* and is not the best of a bad lot, then we should either suspend judgment regarding H_i or accept H_i .

- (12) So, by IBE, we should either suspend judgment regarding the moral reliability of X, Y, and Z or accept their unreliability.(1,9-11)
- (13) Moral realists have no overriding reasons to accept the moral reliability of X, Y, and Z.
- (14) If we should either suspend judgment regarding the moral reliability of X, Y, and Z or accept their unreliability and if moral realists have no overriding reasons to accept the moral reliability of X, Y, and Z, then moral realists should either suspend judgment regarding the moral reliability of X, Y, and Z or accept their unreliability.
- (15) So, moral realists should either suspend judgment regarding the moral reliability of X, Y, and Z or accept their unreliability.(12-14)
- (16) If moral realists should either suspend judgment regarding the moral reliability of X, Y, and Z or accept their unreliability, then moral realists have (inductive) reason for either suspending judgment regarding the moral reliability of the *entire* processes behind our moral beliefs or accepting that these processes are unreliable

- (17) So, moral realists have (inductive) reason for either suspending judgment regarding the moral reliability of the entire processes behind our moral beliefs or accepting that these processes are unreliable. (15-16)
- (18) If moral realists have (inductive) reason for either suspending judgment regarding the moral reliability of the entire processes behind our moral beliefs or accepting that these processes are unreliable, then moral realists have (inductive) reason for thinking we are not justified in holding our moral beliefs.
- (19) So, moral realists have (inductive) reason for thinking we are not justified in holding our moral beliefs. (17, 18)

How might the local version of the challenge work? It might impugn *particular* moral judgments or domains of moral judgment. We could plug in a variety of causal components for X, Y, and Z. For example, we could plug in moral disgust or “repugnance”—a plausible causal source of some prominent intuitions that play an important role in debates in applied ethics—and assess available explanations of its alleged reliability. Though Leon Kass (1997) has suggested that repugnance is wise and reliable, he notably has not offered any explanation for why we should think so and no

plausible realist explanation appears ready to hand. This suggests that disgust-based moral judgments could be very well vulnerable to the local version of the *Explanatory Challenge*.

For another example of how the local version of the challenge could work against a particular domain of moral judgments, consider the case of our demandingness judgments—that is, our judgments that a moral view (e.g. Act Utilitarianism) is mistaken for requiring compliant agents to give up too much (in terms of well-being, central projects, etc). Partiality bias—our evolved disposition to favor one’s interests (or perceived interests) or the interests (or perceived interests) of one’s kin and community—is plausibly a significant causal component of the processes behind our demandingness judgments.²⁵ Since the moral facts are commonsensically impartial in the minimal sense that they do not systematically favor your interests or the interests of your immediate circle over the interests of others, what plausible explanation could there be for how partiality bias is reliably leading us toward commonsensical moral judgments? It is doubtful the realist has one. If this is correct, then, along the lines suggested above, the local version of the *Explanatory Challenge* provides us with some inductive reason for either suspending judgment regarding or denying the reliability of

²⁵ For this case, see Braddock’s “Defusing the Demandingness Objection” (MS).

the *entire* processes behind our demandingness judgments that include the operation of partiality bias as a component.

The more causally significant components of the processes behind our moral judgments we plug in the moral reliability of which the moral realist cannot plausibly explain, the stronger (inductive) reason there is for doubting (denying or suspending judgment regarding) the reliability of the *entire* belief-forming processes that include such components (see premises (16) and (17)).

And here is how a more generalized moral skepticism could be vindicated: the more moral judgments whose reliable formation we are led to doubt by this local challenge, the stronger (inductive) reason there is for accepting a more generalized skepticism about moral reliability that should bother the moral realist, since, as argued earlier in this chapter,²⁶ the accessible unreliability of our moral belief-forming processes has a significant skeptical upshot (which is captured by premises (18) and (19)). This local form of the explanatory challenge, then, can pack a generalized skeptical punch even without being committed to the stronger premise (2) of the original (generalized) *Explanatory Challenge* that claims that our best empirical moral psychology provides us with plausible explanations of our moral beliefs.

²⁶ See pp. 237-38.

5.6 Concluding Remarks

Both the original version and local version of the *Explanatory Challenge* show how empirical investigation of the origins of our moral beliefs can bear on their epistemic status. Whether we are justified in holding our moral beliefs depends on whether we have good reason to doubt their reliability. And whether we have good reason to doubt their reliability depends, as the challenge suggests, on the best explanation of our failure to explain moral reliability. So what is the best explanation of the moral realist's failure to explain moral reliability: that there is no such reliability or that we have little to no grip on the processes behind our moral beliefs (or some other explanation)? The more we find out about the origins of our moral beliefs, the less plausible the *Explanation from Causal Ignorance* becomes and the more it seems that the moral realist's failure to explain moral reliability is best explained by the fact that there is no such reliability. The *Explanatory Challenge* thus shows how empirical work can have important implications for moral epistemology.

Chapter 6: Conclusions and Directions for Future Research

The major conclusions of this dissertation are the following:

- (1) Skeptical implications hinge on the truth of the explanatory premise that our best explanations of our moral judgments do not appeal to nor imply their truth. This is because the explanatory premise implies that our moral belief-forming processes are characterized by a kind of epistemic insensitivity that in turn implies their epistemic unreliability (Chapter 2).
- (2) Street's Darwinian Dilemma and Enoch's Epistemological Challenge are inadequate, due to a central defective inference, but the defect is instructive and usefully guides us to a better explanatory challenge in the genre, namely the *Explanatory Challenge* developed in Chapters 4 and 5. (Chapters 3-5)
- (3) Moral realists do not have a plausible explanation of moral reliability (Chapter 4).
- (4) The premise that realists have no plausible explanation of moral reliability is empirically tractable and supported by our best empirical understanding of the origins of our moral judgments (Chapter 4).
- (5) Skeptical implications for moral realism hinge on the premise that realists have no plausible explanation of moral reliability. This is so because the

best explanation of the fact that realists have no plausible explanation of moral reliability is that there is no such reliability (Chapter 5).

- (6) Even if the *Explanatory Challenge* is unsuccessful, a local version of it can be plausible and skeptically significant (Chapter 5).

Directions for future research include:

- (1) Determine whether moral explanations—explanations of our moral judgments that appeal to the moral facts—are part of the best explanations of our moral judgments. (cf. Chapter 2).
- (2) Determine whether the two new moral debunking challenges developed in the dissertation—the *New Debunking Argument from Insensitivity* and the *Explanatory Challenge*—apply to other interesting domains of judgments, religious judgments and various domains of philosophical judgment, for example. (cf. Chapters 2, 4, 5).
- (3) Determine whether and (if so) which prominent *response-dependent* or “constructivist” metaethical views are undermined by the two new moral debunking challenges developed in the dissertation (cf. Chapters 2, 4, 5).
- (4) Reconstruct and evaluate *alternative* explanations of moral reliability (cf. Chapter 4).

- (5) Assess the moral reliability of other empirically confirmed and causally significant (biological and cultural) evolutionary processes and more proximate psychological causes (e.g. specific emotions) (cf. Chapter 4).
- (6) Determine the skeptical scope of the local version of the *Explanatory Challenge*—that is, whether it can secure a generalized moral skepticism or can at least undermine various domains of moral judgment or particular moral judgments. Accordingly, identify prominent moral judgments in normative ethics, applied ethics, and political philosophy that the local version could undermine (cf. Chapter 5).

References

- Batson, C. Daniel, Klein, T., Highberger, L., and L. Shaw. (1995). "Immorality from Empathy-Induced Altruism: When Compassion and Justice Conflict." *Journal of Personality and Social Psychology* 68(6): 1042-1054.
- Becker, Kelly. (2008). *Epistemology Modalized*. London: Routledge.
- Benacerraf, Paul. (1973). "Mathematical Truth." *The Journal of Philosophy* 70: 661-679.
- Blackburn, Simon. (1971). "Moral Realism." In *Essays in Quasi-Realism*. Oxford: Oxford Univ. Press.
- . (1986). "Supervenience Revisited." In *Essays in Quasi-Realism*. Oxford: Oxford Univ. Press.
- Boyd, R. and Richerson, P. (1992). "Punishment Allows the Evolution of Cooperation (Or Anything Else) in Sizable Groups." *Ethology and Sociobiology* 13: 171-195.
- . (2005). *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: Univ. of Chicago Press.
- Boyer, Pascal. (2001). *Religion Explained*. New York: Basic Books.
- Braddock, Matthew. (MS). "Defusing the Demandingness Objection."
- Brink, David. (1989). *Moral Realism and the Foundations of Ethics*. New York: Cambridge Univ. Press.
- Campbell, Richmond. (1997). "Critical Notice of David Copp's *Morality, Normativity, and Society*." *Canadian Journal of Philosophy* 27(3): 423-444.
- Carruthers, Peter. (2004). *The Nature of the Mind*. London: Routledge.
- Chomsky, Noam. (1957). *Syntactic Structures*. The Hague: Mouton.
- Clarke-Doane, Justin. (2012). "Morality and Mathematics: The Evolutionary Challenge." *Ethics* 122: 313-340.

Copp, David. (2006). "Darwinian Skepticism About Moral Realism." *Philosophical Issues* 18 (1):186-206.

Dawkins, Richard. (2006). *The God Delusion*. New York: Houghton Mifflin.

De Cruz, Helen. (2007). "Innate Ideas as a Naturalistic Source of Mathematical Knowledge: Towards a Darwinian Approach to Mathematics." Dissertation. Vrije Universiteit Brussel.

Enoch, David. (2010). "The Epistemological Challenge to Metanormative Realism." *Philosophical Studies* 48: 413–438.

----. (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford Univ. Press.

Field, Hartry. (1989). *Realism, Mathematics and Modality*. New York: Blackwell.

Fitzpatrick, William. (2008). "Morality and Evolutionary Biology." *Stanford Encyclopedia of Philosophy*.

Gibbard, Allan. (2011). "How Much Realism? Evolved Thinkers and Normative Concepts." In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics* Vol. 6. Oxford: Oxford University Press.

Goldman, Alvin. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard Univ. Press.

----- (1989). "Metaphysics, Mind, and Mental Science." *Philosophical Topics* 17: 131-145.

----- (1992). "Cognition and Modal Metaphysics." In *Liaisons*. Alvin Goldman. Cambridge, MA: MIT Press.

Goodwin, Geoffrey and John Darley. (2008). "The Psychology of Meta-Ethics: Exploring Objectivism." *Cognition* 106: 1339-66.

Haidt, Jonathan. (2007). "The New Synthesis in Moral Psychology." *Science* 316: 998-1002.

- Harman, Gilbert. (1977). *The Nature of Morality*. New York: Oxford Univ. Press.
- (1999). "Moral Philosophy and Linguistics." *Proceedings of the 20th World Congress of Philosophy, Volume 1: Ethic*. Klaus Brinkmann, ed. Bowling Green, Ohio: Philosophy Documentation Center: 107-115.
- Hauser, Marc. (2006). *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: HarperCollins.
- Henderson, David and Terry Horgan. (2007). "Some Ins and Outs of Transglobal Reliabilism." In *Internalism and Externalism in Semantics and Epistemology*. Sanford Goldberg (ed.). Oxford: Oxford Univ. Press, pp. 100-130.
- Hills, Alison. (2009). "Supervenience and Moral Realism." In *Reduction, Abstractions, and Analysis*. Ontos Verlag, pp. 163-177.
- Horsey, Richard. (2002). "The art of chicken sexing." *UCL Working Papers in Linguistics* 14, pp. 107-117.
- Jackson, Frank. (1998). *From Metaphysics to Ethics*. Oxford: Oxford Univ. Press.
- James, Scott M. (2011). *An Introduction to Evolutionary Ethics*. Oxford: Wiley-Blackwell.
- Joyce, Richard. (2001). *The Myth of Morality*. Cambridge: Cambridge Univ. Press.
- (2007). *The Evolution of Morality*, Cambridge, MA: MIT Press.
- Kass, Leon. (1997). "The Wisdom of Repugnance." *New Republic* 216(22): 17-26.
- Kogut, Tehila, and Llana Ritov. (2011). "The Identifiable Victim Effect: Causes and Boundary Conditions." In *The Science of Giving: Experimental Approaches to the Study of Charity*. Eds. Daniel Oppenheimer and Christopher Olivola. New York: Psychology Press.
- Kornblith, Hiliary. (2008). "Knowledge Needs No Justification." In *Epistemology: New Essays*. Quentin Smith (ed.). Oxford: Oxford Univ. Press, pp. 5-23.

- Kramer, Matthew. (2009). *Moral Realism as a Moral Doctrine*. Oxford Univ. Press.
- Leiter, Brian. (2001). "Moral Facts and Best Explanations." *Social Philosophy and Policy* 18: 79–101.
- Levine, M., Cassidy, C., Brazier, G., and S. Reicher. (2002). "Self-categorization and bystander non-intervention: Two experimental studies." *Journal of Applied Social Psychology* 32(7): 1452-1463.
- Levy, Neil. (2004). *What Makes Us Moral? Crossing the Boundaries of Biology*. Oxford: Oneworld.
- Lewis, David. (1973). *Counterfactuals*. Cambridge, MA: Harvard Univ. Press.
- Lipton, Peter. (2004). *Inference to the Best Explanation*. 2nd edition. London: Routledge.
- Luper-Foy, S. (ed.). (1987). *The Possibility of Knowledge: Nozick and His Critics*. Totowa, NJ: Rowman & Littlefield.
- Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin.
- Majors, Brad. (2007). "Moral Explanation." *Philosophy Compass* 2(1): 1-15.
- McGinn, Colin. (1993). *Problems in Philosophy: The Limits of Inquiry*. Oxford: Wiley-Blackwell.
- (1994). "The Problem of Philosophy." *Philosophical Studies* 76: 133-156.
- Meyerson, Denise. (1991). *False Consciousness*. Oxford: Oxford Univ. Press.
- Mikhail, John. (2011). *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. New York: Cambridge Univ. Press.
- Murray, Michael J. (2009). "Scientific Explanations of Religion and the Justification of Religious Belief." In *The Believing Primate*, (eds.) Jeffrey Schloss and Michael J. Murray. Oxford: Oxford Univ. Press.

- Nichols, Shaun. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford Univ. Press.
- Nolan, Daniel. (1997). "Impossible Worlds: A Modest Approach." *Notre Dame Journal of Formal Logic* 38: 535–72.
- Nozick, Robert. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard Univ. Press.
- Parfit, Derek. (1984). *Reasons and Persons*. Oxford: Clarendon.
- Plantinga, Alvin. (1993). *Warrant and Proper Function*. New York: Oxford Univ. Press.
- Pollock, John. (1976). *Subjunctive Reasoning*. Reidel: Dordrecht.
- Prinz, Jesse. (2005). *The Emotional Construction of Morals*. Oxford: Oxford Univ. Press.
- . (2007). "Is Morality Innate?" In *Moral Psychology, Volume 1: The Evolution of Morality*. W. Sinnott-Armstrong, ed. Cambridge, MA: MIT Press. Pp. 367-406.
- . (2011). "Is Empathy Necessary for Morality?" In *Empathy: Philosophical and Psychological Perspectives*, eds. Amy Coplan and Peter Goldie. Oxford: Oxford Univ. Press.
- Pust, Joel. (2001). "Against Explanationist Skepticism Regarding Philosophical Intuitions." *Philosophical Studies* 106(3): 227-258.
- Railton, Peter. (1985). "Moral Realism." *Philosophical Review* 95: 163-207.
- Ridge, Michael. (2007). "Anti-reductionism and Supervenience." *Journal of Moral Philosophy* 4.3: 330-348.
- . (2008). "Moral Non-Naturalism." *Stanford Encyclopedia of Philosophy*.
- Sarkissian, Hagop, John Park, David Tien, Jennifer Wright, and Joshua Knobe. (2011). "Folk Moral Relativism." *Mind & Language* 26(4): 482-505
- Shafer-Landau, Russ. (2001). "Knowing Right from Wrong." *Australasian Journal of Philosophy* 79: 62-80.

- (2003). *Moral Realism: A Defence*. Oxford: Oxford Univ. Press.
- Singer, Peter. (1981). *The Expanding Circle: Ethics and Sociobiology*. New York: Farrfar, Straus, and Giroux.
- (1995). *How Are We to Live: Ethics in an Age of Self-Interest*. Amherst, NY: Prometheus.
- (1999). "A Response." In *Singer and His Critics*, ed. Dale Jamieson. Oxford: Blackwell.
- Sinnott-Armstrong, Walter. (2006). *Moral Skepticisms*. New York: Oxford Univ. Press.
- (2008). "Framing Moral Intuitions." Sinnott-Armstrong (ed.). *Moral Psychology, Vol. 2: The Cognitive Science of Morality*. MIT Press, 47-76.
- Sobel, David. (1998). "Review of David Copp's *Morality, Normativity, and Society*." *Economics and Philosophy* 14(2): 349-357.
- Sosa, Ernest. (2002). "Reliability and the A Priori." In *Conceivability and Possibility*, ed. Tamar Szabo Gendler and John Hawthorne. Oxford: Oxford Univ. Press.
- Sommers, Tamler, and Alex Rosenberg. (2003). "Darwin's Nihilistic Idea: Evolution and the Meaninglessness of Life." *Biology and Philosophy* 18: 653-68.
- Sosa, Ernest. (1999). "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13: 137-49.
- Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*. Cambridge, Mass: Blackwell.
- Sripada, Chandra. (2008). "Nativism and Moral Psychology: Three Models of the Innate Structure That Shapes the Contents of Moral Norms." In *Moral Psychology, Volume 1: The Evolution of Morality*. W. Sinnott-Armstrong, ed. Cambridge, MA: MIT Press. Pp. 319-343.

Stalnaker, Robert. (1968) "A Theory of Conditionals." *Studies in Logical Theory, American Philosophical Quarterly*, Monograph 2: 98-112.

Stein, Edward. (1996). *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science*. Oxford: Clarendon.

Street, Sharon. (2006). "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127: 109-166.

----- (2008). "Reply to Copp." *Philosophical Issues* 18(1): 207-228.

Sturgeon, Nicholas. (1984). "Moral Explanations." In *Morality, Reason and Truth*. (eds.) David Copp and D. Zimmerman. Totowa, NJ: Rowman and Allanheld, 49-78.

Stürmer, S., M. Snyder, A. Omoto. (2005). "Prosocial Emotions and Helping: The Moderating Role of Group Membership." *Journal of Personality and Social Psychology* 88: 532-46.

Vander Laan, David. (2004). "Counterpossibles and Similarity." In *Lewisian Themes*. Frank Jackson and Graham Priest(eds.). Oxford: Oxford University Press.

Van Fraassen, Bas. (1989). *Laws and Symmetry*. Oxford: Clarendon Press.

Vogel, Jonathan. (1990). "Cartesian Skepticism and Inference to the Best Explanation." *Journal of Philosophy* 87: 658-666.

----- (2000). "Reliabilism Leveled." *Journal of Philosophy* 97: 602-623.

Wielenberg, Erik J. (2010). "On the Evolutionary Debunking of Morality." *Ethics* 120: 441-464.

Zagzebski, Linda. (1996). *Virtues of the Mind*. Cambridge: Cambridge Univ. Press.

Biography

Matthew C. Braddock was born on May 30, 1984. He earned his Ph.D in Philosophy from Duke University in 2012. He is currently Assistant Professor of Philosophy at University of Tennessee, Martin. Recent publications include:

- (2012). (with Alex Rosenberg). "Reconstruction in Moral Philosophy?" *Analyse & Kritik* 34(1): 63-80.
- (2010). "Constructivist Experimental Philosophy on Well-Being and Virtue." *Southern Journal of Philosophy*. 48(3): 295-323.
- (2009). "Evolutionary Psychology's Moral Implications." *Biology and Philosophy*. 34(4): 531-540.