Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

8-20-2012 12:00 AM

Does Empirical Moral Psychology Rest on a Mistake? Understanding Theories About the Nature of Moral Judgment as Moral Propositions

Patrick Clipsham, *The University of Western Ontario*

Supervisor: Dr. Anthony Skelton, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Philosophy
© Patrick Clipsham 2012

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Ethics and Political Philosophy Commons

Recommended Citation
Clipsham, Patrick, "Does Empirical Moral Psychology Rest on a Mistake? Understanding Theories About the Nature of Moral Judgment as Moral Propositions" (2012). *Electronic Thesis and Dissertation Repository*. 763.
https://ir.lib.uwo.ca/etd/763

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

DOES EMPIRICAL MORAL PSYCHOLOGY REST ON A MISTAKE? UNDERSTANDING THEORIES ABOUT THE NATURE OF MORAL JUDGMENT AS MORAL PROPOSITIONS

(Spine title: Does Empirical Moral Psychology Rest on a Mistake?)

(Thesis format: Monograph)

by

Patrick Clipsham

Graduate Program in Philosophy

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Patrick Clipsham 2012

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor

Dr. Anthony Skelton

Supervisory Committee

Dr. Andrew Botterell

Dr. Carolyn McLeod

Examiners

Dr. Sergio Tenenbaum

Dr. Andrew Botterell

Dr. Barry Hoffmaster

Dr. Charles Jones

The thesis by

Patrick Clipsham

entitled:

Does Empirical Moral Psychology Rest on a Mistake? Understanding Theories About The Nature of Moral Judgment as Moral Propositions

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date

Chair of the Thesis Examination Board

Abstract

The main goal of this dissertation is to develop and defend the thesis that theories about the nature of moral judgment must be understood as carrying moral commitments. This has profound consequences for the methodology of metaethics. Specifically, it implies that theories about the nature of moral judgment cannot be understood as empirical hypotheses.

There have historically been many attempts to develop a philosophically satisfying theory that characterizes the nature and content of moral judgments. Many philosophers have thought that such theories are best understood as morally neutral hypotheses about human psychology. Recently, a number of philosophers have attempted to approach this question by treating theories about the nature of moral judgment as empirical hypotheses that can be confirmed or disconfirmed by psychological and neuroscientific evidence. I argue that this methodological presupposition is mistaken.

In the first and second chapter, I articulate and defend a test for identifying moral propositions and use it to demonstrate that a number of prominent metaethicists have mistakenly thought that theories about the nature of moral judgment are morally neutral. The third chapter begins with an argument that moral propositions cannot be identical to or definable in terms of empirically-confirmable hypotheses. I first survey a number of prevalent arguments in favour of the claim that moral propositions must be distinguished from empirically-confirmable hypotheses. I conclude that very few of these arguments are convincing, but I expand upon a version of one of the arguments (the so-called "What's at issue?" argument) and demonstrate that it succeeds where others fail. I then show that this conclusion undermines several empirically-informed theories of moral judgment put forward by Shaun Nichols, Jesse Prinz, and Richard Joyce.

I close by arguing that many considerations may be relevant to the confirmation or disconfirmation of theories of moral judgments, including empirical evidence. However, the question of whether or not such considerations are relevant is itself a moral

question. Because of this feature of metaethical inquiry, theories about the nature of moral judgment must always be predicated on substantial moral assumptions and thus, at bottom, must always be understood as having moral content

Keywords

Metaethics, Normative Ethics, Moral Psychology, Empirical Hypotheses, Moral Judgment

Acknowledgments

I would like to thank everyone who supported me while writing this thesis. I benefitted greatly from advice and feedback given to me by my supervisory committee members Dr. Carolyn McLeod and Dr. Andrew Botterell, as well as all the members of my examination committee. I would also like to thank Dr. Allan Gibbard for sponsoring my visit to the University of Michigan in the winter term of 2011 and for discussing my ideas with me at length, and Katy Fulfer for reading earlier drafts of this thesis and offering helpful advice on numerous occasions. Most importantly, I would like to thank Dr. Anthony Skelton for reading many drafts of this document over the past two years and engaging me in countless helpful discussions about the content of this dissertation. Finally, thank you to my friends for all of their support, and my parents for everything they have done for me.

Table of Contents

CERTIFICATE OF EXAMINATION....................ii

Abstract....................iii

Acknowledgments....................v

Table of Contents....................vi

Introduction - Metaethical Neutrality: Then and Now....................1

Chapter One – Differentiating the Moral and the Non-Moral....................16

I – Preliminaries....................19

I.i – Propositions, Content, and Methodological Assumptions....................19

I.ii – A Looming Objection....................24

I.iii – Introducing My Proposal: MORAL....................26

II –Alternative Theories....................33

III – Virtues of and Problems for MORAL....................46

III.i – Mixed Compounds....................47

III.ii – The Reductio *and The Proviso*....................48

III.iii – Converse Entailment and Moral Advice....................52

IV - Conclusion....................55

Chapter Two – Theories of Moral Judgment as Normative Moral Theories....................57

I – Mackie and the Claim to Objectivity....................58

II- Relativism, Ideal Observer Theory, and Neo-Sentimentalism....................63

III - Objections....................74

III.i – Anthropology or Morality?....................74

III.ii – Consistency with Normative Theories....................76

III.iii – Implications for Action?....................77

IV- Conclusion....................92

Chapter Three – Theories of Moral Judgment are not Empirical Hypotheses.........93
I – Empirical Hypotheses...94
II – Non-Naturalism and Ethics as a Branch of Philosophy................................97
III- Non-Natural Concepts and Moorean Considerations....................................100
IV – Morality, Convention, and Nichols's Sentimental Rules Account.............120
V – Prinz's Sentimentalist Relativism...129
VI – Joyce's Minimal Projectivism...146
VII – Psychological Realism...155
VIII – Moral Methods...167

Conclusion..175

Bibliography..179

Curriculum Vitae..185

Introduction

Metaethical Neutrality: Then and Now

The central thesis that I will defend in this dissertation is that theories about the psychological nature of moral judgment must be construed as being internal to moral discourse. This conclusion may be counterintuitive for a number of reasons. First, my thesis blurs the line between normative ethical questions and supposedly higher order, metaethical questions. Second, it implies that the findings of empirical sciences cannot provide us with an answer to the question 'what are moral judgments?' Third, it requires theories of moral judgment to be confirmed or disconfirmed by moral considerations. My arguments in favour of this thesis and my discussion of these potentially counter-intuitive conclusions fill the bulk of the pages that follow. Before I introduce these arguments, we require certain preliminaries. The most pressing one is: what are theories about the nature of moral judgment, and why are they important?

Theories about the psychological nature of moral judgment serve as a crucial starting point for a number of metaethical and normative ethical discussions. These discussions are complicated because there are many questions that are relevant to figuring out what, exactly, moral judgments are. Are moral judgments primarily derived from reason? Do emotions make a significant contribution to moral judgment? What kind of role do emotions play? Are there any distinctive features of moral judgments that can be used to differentiate them from non-moral judgments? What are the objects of moral

judgments? These are questions that have long been at the forefront of philosophical research on the nature of moral judgment.

In recent decades, there has been a strong tendency for philosophers to treat these and other related questions as empirical hypotheses. While many philosophers have approached these philosophical questions from an empirical perspective, three philosophers are particularly explicit about this goal: Shaun Nichols, Jesse Prinz, and Richard Joyce. Nichols, for example, encourages philosophers engaged with these questions to pay attention to “controlled experiments in moral psychology” in order to resolve debates about “the nature of moral judgment ranged widely, including issues about the role of sentiment in moral judgment, the role of reason in moral judgment, and the origin of moral judgment” (Nichols 2004 p. 4).

In addition to developing a detailed account of the evolution of moral judgment,[1] Richard Joyce has also argued that a number of theories about the nature of moral judgment should be refined into empirically testable hypotheses. His specific focus is on the theory known as moral projectivism according to which the seemingly objective nature of moral properties is explained by a metaphorical projection of one’s attitudes onto an otherwise value-neutral world. Joyce says of moral projectivism that “it is an empirical hypothesis and thus must be tested as such” (Joyce 2009 p. 53).

Jesse Prinz endorses a philosophical method according to which “philosophers cite laboratory studies in support of their theories” (Prinz 2007b p. 271). Prinz appeals to such studies in order to defend a theory of moral judgment that attributes to emotions a

[1] See Joyce 2006

large role in moral judgment and understands the truth conditions of moral judgments as being relative to individuals (a moral fact may be true for an individual with one sentimental disposition and false for another individual with different dispositions). Further, he believes that these empirically-supported theories have epistemic priority over philosophical theories about the nature of moral judgment that lack empirical support: "here, as elsewhere, we are hostage to empirical fortune, and I think relativism is supported by the evidence" (Prinz 2007b p. 285).

These three philosophers are unified in two respects. First, they all endorse and promote an empirical approach to many difficult philosophical questions. Second, they all attempt to offer an empirically-supported answer to one such question. Specifically, all of these philosophers articulate and defend *philosophical theories about the nature and content of moral judgment.* There are three important features of this type of theory.

First, these theories are *philosophical* insofar as they rival and contradict theories that philosophers have historically defended about the nature of morality and moral judgment. Rather than attempting to offer a rigorous definition of a 'philosophical theory', I will simply enumerate some examples. One example of such a philosophical theory can be seen in the debate between cognitivism and non-cognitivism. Cognitivist theories hold that moral judgments express beliefs which are capable of being true or false (Shafer-Landau 2003 p. 17). Non-cognitivism, however, states that moral judgments do not express beliefs, but rather express emotive states (Ayer 1936).[2] Any theory that

[2] There are actually many forms of non-cognitivism with varying degrees of complexity. Not all such theories are this simplistic. Stevenson's emotivism, for example, understands a moral judgment as having both a descriptive and an imperatival component (Stevenson, 1937). I focus here on A.J. Ayer's version because it helps illuminate what is at issue in philosophical debates about the nature of moral judgment.

purports to confirm or disconfirm cognitivism or non-cognitivism would therefore be philosophical in my sense of the term.

Another example of a philosophical disagreement about the nature of moral judgment is that between motivational externalists and motivational internalists. One common understanding of motivational internalism amounts to the claim that whenever an individual recognizes a moral obligation to do X, this person will be motivated to do X. Externalist theories deny the truth of internalism (see Brink 1989 p. 38). Any theory that deals with the relationship between motivation and moral judgment is a philosophical theory. These examples do not exhaust the philosophical topics that pertain to the nature of moral judgment, but they illustrate the kinds of questions philosophical theories of moral judgment address. Nichols, Prinz, and Joyce all advocate the use of empirical methods in order to work towards the resolution of these and similar debates. Their theories are therefore philosophical.

Second, the theories put forward by Nichols, Prinz and Joyce are all about the *nature* of moral judgment. I understand this phrase in a fairly broad way. There are a number of topics that are relevant to determining the nature of moral judgment, including the role of emotions in moral judgment, the role of reason in moral judgment, the distinction between moral judgments and non-moral judgments, and the relationship between moral judgment and motivation. I address many theories of the nature of moral judgment in the chapters that follow, and one of the major unifying features of these theories is that they are all understood by their proponents as offering answers to the question 'what are moral judgments?' A short list of theories I discuss that offer partial answers to this question are: relativism, sentimentalism, neo-sentimentalism, ideal-

observer theory and the thesis that moral judgments understand moral properties as objective features of the world.[3] While there is a sense according to which many of these theories are asking very different questions (about say, the origins of moral judgment rather than the phenomenology of moral judgment), they are all contributing at least partial answers to the question 'what are moral judgments?'

Some of the theories I discuss do more than make abstract claims about the nature of moral judgment. Some of them go so far as to make claims about what the *object* of moral judgments are or, in other words, what moral judgments are *about*. This brings me to the third important feature of many of the theories I will be discussing: many of them make claims about the *content* of moral judgments. A theory about the content of moral judgments takes a stand on questions about the subject matter of this class of judgments. For example, many forms of relativism claim that moral judgments are, in fact, about what sentiments an individual (or an individual's community) has towards a certain action (see Prinz 2007a). Other theories claim that moral judgments are not about what sentiments certain individuals *do* have but rather what emotions individuals *should* have or, more precisely, what emotions it is rational to have (see Gibbard 1990). These are rival theories about the content of moral judgments insofar as they make inconsistent claims about what moral judgments are about or what the objects of these judgments are. Questions about the nature of moral judgment and questions about the content of moral judgments substantially overlap, but nonetheless can be distinguished.

The following chapters are primarily concerned with philosophical theories about the nature and content of moral judgment. When I refer to 'theories of moral judgment'

---

[3] The details of these theories will be outlined in later chapters.

or state that a theory is about the nature of moral judgment, I should be understood as claiming that the theory under discussion is a philosophical theory about the nature and/or content of moral judgments.

The argument I develop in the following chapters leads to the conclusion that many philosophical theories about the nature and content of moral judgment must be understood as first-order moral claims. This implies that the proper methodology to employ in order to decide between these theories is the same methodology that we should use to decide between normative moral theories and to resolve other moral disagreements. Further, I argue that this feature of theories about the nature and content of moral judgment implies that they cannot be construed as empirical hypotheses. Philosophers such as Nichols, Prinz and Joyce who have attempted to use empirical methods in order to defend a philosophical theory of the nature of moral judgment have therefore unknowingly committed a methodological mistake.

This methodological error is not unique to the individual philosophers I will be discussing below, and I am not the first to criticize it. The assumptions these philosophers make about the moral neutrality of the theories they defend mirrors the assumptions many philosophers have endorsed about the moral neutrality of metaethical inquiry.

Philosophers first introduced the idea that metaethical inquiry was morally neutral in the middle of the twentieth century. One of the earliest appeals of this sort was in defense of a form of non-cognitivism. Richard Brandt once argued that C.L. Stevenson's emotivism identifies the wrong class of reasons as being relevant to morality (Brandt 1950 pp. 312-313). Brandt's argument amounted to the claim that Stevenson's emotivism

entails some counter-intuitive moral propositions about what sorts of factors are relevant to determining whether or not a particular individual has a moral reason to act in a certain way. Stevenson retorted that his project did not purport to establish any moral propositions. His theory was supposed to be understood as an attempt to "understand what goes on in ethics" and he characterized this metanormative inquiry as "being itself nonnormative" (Stevenson 1950 p. 528). In his reply, Stevenson was one of the first philosophers to explicitly appeal to the moral neutrality of metaethics.[4] William Frankena subsequently discussed the distinction between normative ethics and metaethics in the following terms:

> …it will be convenient to divide ethics into two parts: (1) ethical theory or metaethics, which asks such questions as what is the meaning of the terms 'right,' 'good,' etc., and consists not of ethical judgments proper, but of such logical, epistemological, or ontological statements as 'good means desired,' 'right stands for a non-natural property,' 'ought implies can'; (2) normative ethics, which asks what things or actions are good, right, etc., and consists of ethical judgments (i.e., judgments of value and obligation) proper. (Frankena 1951 pp. 44-55.)

According to this understanding, only the latter must be understood as being morally committal, as the former is morally neutral. This way of thinking about metaethics has informed much of moral philosophy in the last sixty years.

[4] For the sake of clarity, I will *not* be arguing that metaethical inquiry is normative, nor will I be arguing that empirical investigations are normatively neutral. Rather than focusing on the distinction between the normative and the non-normative, I will be discussing the distinction between the moral and the non-moral. I will be assuming that any proposition that is moral will also be normative, but that not all propositions which are normative are moral.

There is, however, also a long tradition of philosophers claiming that so-called "metaethical" theses actually carry moral commitments. One of the first proponents of this view writes that "the distinction between 'normative' and 'meta-ethical' theories is not at all clear, that by common-sense standards contemporary theories *are* normative (specifically, moral) in character, and that a more rigorous distinction is apt to involve circularity." (Mothersill 1952 pp. 587-594). Bernard Williams has expressed a similar attitude:

> The distinction between the ethical and meta-ethical is no longer found to be so convincing or important. There are several reasons for this, but the most relevant here is that it is now obvious…that what one thinks about the subject matter of ethical thought, what one supposes it to be about, must itself affect what tests for acceptability or coherence are appropriate to it; and the use of those tests must affect any substantive ethical results. (Williams 1985 p. 73)

Many other philosophers have argued for related theses.[5] Ronald Dworkin has more recently articulated an influential version of this argument. His goal was to undermine skeptical metaethical theories by showing that there is no intelligible way of engaging in metaethical inquiry outside of first-order ethics:

> Any successful – really, any intelligible – argument that evaluative propositions are neither true nor false must be internal to the evaluative domain rather than archimedean about it. So, for example, the thesis that there is no right answer to the question whether abortion is wicked is itself a substantive moral claim, which must be judged and evaluated in the same way as any other substantive moral claim. (Dworkin 1996 p. 89)

[5] See, for example, Olafson (1956), Taylor (1958), and Gewirth (1960)

The general idea that Williams, Mothersill, Dworkin and others defend is that metaethical theories *about* morality and normative moral theories that are internal to the moral domain are much more closely related than many philosophers have thought. In many cases, these theories seem to inform and justify one another, and the distinction between them seems to be a matter of degree rather than a matter of kind. Because of these kinds of considerations, a number of philosophers have concluded that metaethical theories cannot be understood as being morally neutral. This view is sometimes expressed as the claim that metaethics is normative, or that genuine metaethical inquiry is impossible.

It is not, in my view, useful to make broad pronouncements about the status and nature of metaethics as a whole. The term 'metaethics' is sufficiently vague that projects of this kind must rely on a prior definition of it. By presuming such a definition, however, one risks assuming a characterization of metaethics at the outset that begs the question against one's opponents.

In short, I agree with L.W. Sumner that "one can…draw the conclusion that metaethics is an incoherent notion and hence impossible, but doing so solves no important problems and amounts simply to announcing how one plans to talk" (Sumner 1967 p. 105). Rather than making grandiose claims about the status of metaethical inquiry as a whole, I will focus my discussion on particular theories with the specific subject matter described above. There are nonetheless many commonalities between my arguments and the arguments put forward by philosophers such as Mothersill, Dworkin, and Williams. Like all three of them, I will conclude that a class of theories which are generally thought to be neutral, austere, and metaethical must be understood as normative

moral theories. Furthermore, I will conclude that this fact has a substantial impact on the method we should use to decide between these theories. In particular, I will argue that theories about the nature of moral judgment must be confirmed or disconfirmed by the same method we use to confirm or disconfirm moral propositions.

This argument is broken down into three chapters. In the first chapter, I argue in favour of a test for determining whether or not a propositions is moral. Many philosophers have proposed tests for differentiating moral from non-moral propositions (I focus on three: Karmo 1988, Dreier 2002, Fantl 2006), but I argue that they all succumb to a common criticism. Specifically, I argue that they are all committed to the erroneous assumption that any proposition that is logically inconsistent (in classical logic) with a moral proposition must be understood as a moral proposition.[6] This commitment is widely held and is largely indebted to a remark made by David Hume. I demonstrate that this commitment implies that many paradigmatically morally-neutral propositions (such as, for example, the proposition ‘some peas are green’) are actually moral propositions.

My alternative proposal abandons the idea that any proposition that is logically inconsistent with a moral proposition has moral content (or, in other words, carries a moral commitment) and emphasizes instead the possible contribution that a proposition (not the propositions it entails) can make to moral discourse. I articulate this criterion as the claim that if a proposition can be used in order to ascribe a moral predicate to an act or state of affairs (or, alternatively, can be used to give moral advice, or add moral

[6] Karmo’s view differs slightly from the others, but I argue in Chapter One that this difference is not important for the purposes of my argument.

information in specific circumstances), the proposition must be understood as having moral content.

In the second chapter, I employ this theory of the nature of moral content in order to demonstrate that a number of philosophers have mistakenly believed that theories about the nature of moral judgment can be rigidly distinguished from first-order, normative moral theses. The goal of this chapter is to show that the methodological presupposition I have attributed to Nichols, Prinz and Joyce is not unique to them, but has actually been held by a large number of highly influential metaethicists. Specifically, I argue that J.L. Mackie's assumption that moral judgments include a claim to objectivity must be understood as a moral proposition despite the fact that Mackie explicitly believes otherwise (Mackie 1977). I then argue for the same conclusion regarding Gilbert Harman's moral relativism (Harman 1975), Michael Smith's ideal-observer theory (Smith 1994) and Allan Gibbard's neo-sentimentalism (Gibbard 1990). All four of these philosophers believe that they are offering morally-neutral accounts of the nature and content of moral judgments, yet all four wind up defending moral propositions that could be confirmed or disconfirmed only by moral evidence.

I then conclude the second chapter by responding to three plausible criticisms of my claim that the theories defended by these philosophers have moral content. First, I address the argument that these theories are best understood as anthropological claims, rather than moral ones. I respond to this by arguing that some anthropological claims about the moral systems used by particular societies are morally neutral, but that philosophers often take these anthropological theories to imply claims about the nature of

moral judgment *simpliciter*. It is this latter class of propositions that I argue carry moral commitments and thus cannot be confirmed or disconfirmed by anthropological evidence.

Second, I consider the argument that the theories discussed above are morally neutral because they are consistent with a diverse number of moral propositions and normative moral theories. I respond that this may be true, but that this conclusion is not sufficient to show that the theories discussed above lack moral content. Many paradigmatically moral propositions are neutral with respect to a diverse number of normative moral theories, but this does not disqualify them from being moral propositions.

Finally, I discuss an objection that is predicated on the claim that it is a conceptual truth that first-order moral propositions have implications for action, but that this is not the case for the theories discussed above. This objection capitalizes on the fact that it seems to be incoherent to deny that some moral propositions give us reasons to act in certain ways, yet it seems coherent to deny that theories about the nature of moral judgment give us such reasons. This is a complex argument, and it merits an in-depth response. I conclude that the argument fails because it is not a conceptual truth that moral propositions have implications for action.

I begin the third and final chapter with an argument that moral propositions cannot be understood as identical to or reducible to any empirically-confirmable hypothesis. (I follow Nichols, Joyce, Prinz and others in explicating the term 'empirically-confirmable hypothesis' by appealing to a disciplinarian definition.) I survey several arguments that are indebted to G.E. Moore's discussion of the claim that a

number of philosophers have committed the so-called “naturalistic fallacy.” Despite the fact that contemporary philosophers have rejected many of these arguments, some versions of Moore’s argument that moral propositions cannot be reduced to empirically-confirmable hypotheses remain persuasive. I follow Gibbard (2003) in thinking that the most persuasive of these arguments is implicit in some of the examples that Moore uses to illustrate his point (Gibbard dubs this the ‘What’s at issue?’ argument). Specifically, I argue that distinctively moral disagreements (moral disagreements that cannot be explained as disagreements about the meanings of terms, empirical observations or the norms of belief formation) are meaningful insofar as they track in conversational contexts or, in other words, do not cause bafflement in these contexts. The fact that distinctively moral disagreements are meaningful implies that moral disagreements cannot be reduced to non-moral disagreements. This conclusion in turn implies that moral propositions cannot be identical or reducible to empirically-confirmable hypotheses.

Once this argument is concluded, I turn to Nichols, Prinz and Joyce. I begin by discussing Nichols’ claim that his theory of the nature of moral judgment (a form of sentimentalism) is confirmed by empirical evidence. I demonstrate that Nichols relies very heavily on research performed by Elliot Turiel, and show that this research is predicated on some substantive moral assumptions that Nichols’ opponents would deny. In short, I argue that Nichols is wrong to think that his theory is confirmed by empirical evidence as it is actually grounded on several question-begging, morally-loaded assumptions.

I then argue that Prinz’s relativist account of the nature of moral judgment must be understood as a moral theory. Furthermore, Prinz acknowledges this fact. He explicitly

rejects the conclusion of the Moorean arguments discussed above and he believes that the fact that his theory is moral does not disqualify it from being empirically-confirmable. I argue that, like Nichols, Prinz does not recognize the fact that his theory is grounded on substantive moral assumptions. This, I claim, implies that his theory is not justified by empirical evidence, but rather by his moral assumptions. Furthermore, I show that his moral assumptions are much more counter-intuitive than those of his opponents, and thus should be rejected.

Finally, I discuss Joyce's thesis that moral projectivism is an empirical hypothesis. I argue that Joyce's statement of projectivism disqualifies a certain class of properties from being morally relevant and that the consequence of this is that projectivism rules out a number of possible normative moral theories. Moral projectivism must be understood as a moral proposition that cannot be confirmed or disconfirmed by empirical evidence.

I close this chapter with an argument in favour of the claim that we must predicate our metaethical inquiries on a recognition of the fact that many of these inquiries are moral in character and must be understood as being internal to the normative moral domain. Specifically, I argue that this is true of philosophical theories about the nature and content of moral judgment. Insofar as empirical evidence cannot be appealed to in order to confirm or disconfirm moral propositions, the same must be true for theories about the nature of moral judgment. We must use the same method to confirm or disconfirm these theories as we would use to confirm or disconfirm any other moral proposition. I survey some possible accounts of theory choice in normative ethics and

defend a version of wide reflective equilibrium similar to that endorsed by Norman Daniels (Daniels 1979).

To summarize, I will argue that philosophical theories about the nature and content of moral judgment are moral propositions and that any inquiry into their status must be predicated on a recognition of this fact. I begin by defending a particular account of what it means for a proposition to have moral content. I then demonstrate that some philosophers have misunderstood the distinction between normative ethics and metaethics insofar as they have failed to notice that a class of putatively morally neutral metaethical questions actually possess moral content. Finally, I argue that empirical evidence from natural and social sciences has no bearing on the question 'what is moral judgment?'

Chapter One

Differentiating The Moral and The Non-Moral

The project I described above frequently mentioned the distinction between the moral and non-moral. This distinction is widely invoked and has played a large role in a number of metaethical debates. Much recent metaethical work relies on a conception of the distinction in order to defend particular views about the status of moral realism. David Enoch, for example, offers an explicitly normative moral argument against non-objectivist metaethical theories (Enoch 2011). Simon Blackburn has also argued that the claim that morality is dependent upon human sentiments can only be construed as one that is internal to the moral domain. The consequence of this argument is that theories relying on this so-called "dependency thesis" can be rejected on the basis of the fact that they are committed to a false moral proposition (Blackburn 2010). These arguments attempt to establish that some metaethical claims about the nature of morality must be construed as first-order moral claims. They therefore rely implicitly on a conception of what differentiates the moral from the non-moral.

The distinction between the moral and the non-moral is also invoked in debates about Ronald Dworkin's claim that much of metaethics must be understood as being internal to the moral domain. Tristam McPherson, for example, argues that "the apparent tension between metaethics and the autonomy of morality can be dissolved…by defending a metaethical theory that explains why non-moral theses are irrelevant to normative ethical theorizing" (McPherson 2008 p. 14). According to McPherson, it is possible for there to be informative metaethical theories that are morally neutral, but such

theories must explain and vindicate the autonomy of morality (or, in other words, they explain why it seems to be the case that non-moral considerations do not bear on moral matters.) McPherson takes any such theory to be an adequate counterexample to Dworkin's thesis. Kenneth Ehrenberg has also recently argued that the propositions Dworkin claims are only comprehensible as internal moral propositions can be understood as external, morally-neutral theses (Ehrenberg 2008).

It seems obvious that these debates cannot proceed without an account of what differentiates the moral from the non-moral. And yet most of the individuals engaged in these and other related debates refrain from offering such an account. Some of them go as far as to suggest that it is not possible to draw a rigorous distinction between the two, and are satisfied that our implicit knowledge of the distinction is sufficient for these arguments. This appears to be Dworkin's view when he states the following:

> I shall assume that all readers, including those drawn to Archimedean skepticism, accept that our shared language and common experience include assessments on what we take to be a distinct moral dimension. I shall not attempt to define that dimension, or to separate the predicates we use to employ it. If I am right, no helpful definition of morality as a whole can be given. (Dworkin 1996 pp. 89-90)

Dworkin may be correct that no helpful definition of the moral dimension is forthcoming. Nonetheless, a number of philosophers who have attempted to respond to Dworkin felt that any such response must be predicated on an adequate test for determining whether or not a specific theory is moral. This approach has been pursued most explicitly by James Dreier and Jeremy Fantl. Dreier concludes that certain kinds of actualized secondary-

quality theories are both genuinely metaethical and normatively neutral (Dreier 2002). Fantl, however, uses Dreier's methodology in order to demonstrate the impossibility of engaging in normatively neutral metaethics (Fantl 2006).

I am sympathetic to Dreier's and Fantl's ambitions, as I cannot imagine how it could be possible to engage in a debate about whether or not a set of propositions are morally committal without endorsing at least a provisional theory about what is sufficient evidence for concluding that something is morally committal. I fear, however, that there may be some truth in Dworkin's claim that no helpful definition of morality is forthcoming. I, at least, am not aware of any theory in the philosophical literature that successfully demarcates the moral from the non-moral. The problems for the views I will address below are a testament to how difficult a philosophical issue this is. Nonetheless, before I can discuss whether or not any given proposition is moral (as I will do for much of the remainder of this dissertation), I must clarify how I understand the distinction between moral and non-moral propositions.

In the following section of this chapter, I advocate shifting our focus away from the search for a *definition* of the moral domain, and focusing instead on clearly articulating plausible sufficient conditions which are satisfied only by moral propositions. Whereas Dworkin did not theorize *enough* about the distinction between the moral and the non-moral (as he did not theorize about it at all), I will argue that Karmo, Dreier, and Fantl theorize too much. Their attempts to provide definitions of the moral domain opens them up to a number of objections. Nonetheless, we cannot address these issues without *some* theorizing about the distinction between the moral and the non-moral. It is for this reason that I endorse the approach of seeking sufficient, rather than necessary *and*

sufficient conditions for being moral. My approach therefore commits me to some theoretical claims about the distinction between the moral and the non-moral, but these claims are substantially weaker than those put forward by Karmo, Drier, and Fantl.

In the following section, I explain how I will be using certain terms and explicitly state some of my methodological assumptions. I then defend a particular sufficient condition for having moral content that can serve as a test for determining whether a proposition is moral, and I show why it is superior to the proposals other philosophers have put forward.

I – Preliminaries

I.i –Propositions, Content, and Methodological Assumptions

In this section, I explicitly state how I will be using certain terms and outline some basic methodological assumptions I share with a number of philosophers. As a first step, I must clarify what sorts of things I take to be candidates for having moral content. It is surely plausible to claim that sentences, judgments, assertions, thoughts, theories, beliefs, and many other types of things could be understood as having moral content. In order to eliminate ambiguities regarding what sorts of entities I am considering as candidates for being moral, I will be assuming that when we speak of something having moral content, carrying a moral commitment, being moral, or being internal to the moral domain, we are speaking about the properties of *propositions*. I will be using a very minimal conception of a proposition. I agree with Horwich that an understanding of

propositions as "the things that are believed, stated, supposed, etc; the contents of such states" (Horwich 1990 p. 17) is fairly unproblematic and can be appealed to without engaging in debates about the nature of propositions. I am therefore not endorsing any philosophically contentious theories about the nature of propositions, but am rather using the term as a placeholder for whatever abstract entities stand in the relation described above to beliefs, sentences, judgments, etc. Thinking of propositions in this simplified way is useful because it allows us to discuss moral content in the abstract without focusing exclusively on any one of these mental and linguistic entities. We can therefore understand sentences/theories that express moral propositions and beliefs/judgments/thoughts that have as their object a moral proposition as themselves having moral content. I am willing to be loose with my understanding of moral content, as I am open to the possibility that any of the entities I mentioned above (and maybe several others) could potentially be understood as carrying moral commitments. However, I will not be developing an account of the relationship between propositions and these other potential bearers of moral content. For the sake of simplicity, I will limit myself to discussing only propositions.

There is, however, something peculiar about speaking of propositions as *having* moral content, as they are often thought of as the contents of mental states, or the semantic contents of sentences. Robert Stalnaker, for example, defines a proposition as the "content of an assertion or belief" (Stalnaker 1979 p. 316). This usage is echoed in the quote from Horwich above. While I am sensitive to the fact that my choice of language on this point is peculiar and perhaps even somewhat misleading, I will nonetheless frequently speak of propositions as having moral content. This is not intended to be a

metaphysically loaded notion, however. Rather, for the purposes of this discussion (which is about how best to divide propositions into the categories of moral and non-moral), when I say of a proposition that it has moral content, I am saying nothing more than this: that the proposition in question belongs in the 'moral' category. While there are other options on how to speak about this, they all seem problematic to me. It seems odd, for example, to claim that propositions carry moral commitments, since some philosophers think of acts of assertion as carrying commitments (see, for example, Dreier 2002 p. 248). The phrase 'X is internal to the moral domain' may be more accurate than these other options, but it is far too cumbersome. The simplest option, 'X is moral' is both straightforward and does not employ philosophically contentious terminology. Unfortunately, it is somewhat misleading. To my ear, this phrase seems to be more suited to commending X than it is to identifying X as a proposition that must be understood as being internal to the moral domain. All of these constructions therefore have their downside, and I can see no definitive reason to prefer one over another. I will therefore use them interchangeably, depending on the structure of the sentence in question. Due to its versatility, however, I will primarily use the phrase 'X has moral content' to say of a proposition (X) that it belongs in the 'moral' category.

Much of the philosophical literature I am drawing on assumes that straight ascriptions of moral predicates to acts or states of affairs are uncontroversial and paradigmatic examples of propositions with moral content.[7] I will follow this trend in

[7] Karmo is the first to make such an assumption: "Call a sentence 'uncontroversially ethical' just in case all parties to the logical-autonomy-of-ethics debate would unite in calling it ethical. (There surely are sentences of this kind, for example, 'It ought to be the case that all New Zealanders are shot.' 'Everything that Alfie says is true' and 'Either tea-drinking is common in England or it ought to be the case that all New Zealanders are shot,' on the other hand, are presumably not sentences of this kind: for agreement is presumably lacking on their status)" (Karmo 1988, p. 254). A number of philosophers have followed

assuming that all propositions of the form 'X is morally permissible/impermissible' are paradigmatic examples of moral propositions. The same is true for any straight ascription of other moral predicates, for example, morally good/bad/right/wrong/vicious/virtuous etc. I will assume that there is a finite list of predicates that are morally inflected. In short, I am assuming, along with many other philosophers, that a proposition P will count as a moral proposition – or, as I will sometimes say, will have moral content – if P ascribes a moral predicate to an act or state of affairs.

This account may seem circular, since my theory regarding moral content relies on a prior identification of the nature of moral predicates. However, I believe that the charge of circularity is unfounded. I say this for the following reason: the goal of all the philosophers engaged in this debate is to defend a criterion that can resolve debates about whether or not a specific disputed proposition has moral content. These discussions therefore revolve around a relatively small set of disputed cases, as it is not obviously clear how to classify the propositions in this set. (Some examples of disputed cases are compound propositions that are composed of moral and non-moral atomic sentences, and identity claims that identify moral properties with non-moral properties.) All the philosophers I will discuss begin with the assumption that there is rather large set of undisputed cases, and equate this set with the set of straight ascriptions of moral predicates. The goal of this chapter is to defend a plausible sufficient condition for a proposition's having moral content that can serve as a tool for resolving disagreements

---

Karmo's lead and began their investigations with similar assumptions. Dreier, for example, states that he will "assume that basic, paradigmatically moral statements are predications of moral predicates, especially 'wrong', 'right', 'morally permissible', 'evil', 'good'" (Dreier 2002, p. 244). Fantl follows Dreier and Karmo in these assumptions: "I follow Dreier (2002) in equating the 'class of uncontroversially moral sentences'…with 'the class of atomic sentences that are predications of moral predicates'" (Fantl 2006 p. 25).

about the disputed cases. This account is not meant to *prove* that propositions involving straight ascriptions of moral predicates have moral content, but is, like all the other philosophers who have addressed this question, simply *assuming* that they do as the first step towards figuring out how to resolve debates over the disputed cases. It is therefore not circular to assume that all straight ascriptions of moral predicates have moral content, as I am not setting out to prove this. At the very least, if there is a problem of circularity here, it is not a problem for my account alone, as all alternative accounts make similar assumptions at the outset of their investigations.

It is also not essential for my purposes to offer an exhaustive list of these predicates. All I aim to do is point to the contrast between moral and non-moral predicates by indicating the kinds of predicates I believe belong on this finite list of so-called moral predicates. If an individual believed that different predicates belong on this list, she need only substitute her list in place of my own for the remainder of this argument. All I require is that the reader accept the following two statements: 1) some predicates are moral predicates, and propositions that ascribe these predicates have moral content, and 2) straight ascriptions of predicates that obviously are not moral (such as colour predicates, for example) lack moral content. Given these assumptions, ‘some peas are green’ lacks moral content, whereas ‘all abortions are morally wrong’[8] has moral content. This latter proposition can be thought of as an atomic moral proposition, since it is not built up out of other moral propositions. The class of atomic moral propositions are uncontroversially moral, and propositions like ‘some peas are green’ are

[8] Note that I am here using ‘morally wrong’ as an exemplar of a paradigmatic moral predicate. One could substitute this predicate with any other moral predicate (such as right, good, bad, vicious, virtuous, etc.) without altering my argument.

uncontroversially non-moral. The question under discussion in this chapter is, 'how do we fill in the grey areas between the uncontroversially moral and non-moral propositions?' The assumptions I have made are in service of my attempt to provide a provisional answer to this question.

I.ii – A Looming Objection

One possible way of rejecting my assumptions is to claim that not all propositions of the form 'X is morally permissible/wrong/required/vicious, etc.' have moral content. For example, some philosophers have claimed that understanding moral concepts requires grasping the correct moral theory. I believe that this is Philippa Foot's point when she states that "morality is necessarily connected with such things as justice and the common good, and it is a conceptual matter that this is so" (Foot 1978 p. 92). According to such a theory, any ascription of a moral predicate to things that have something to do with justice and the common good would have moral content, whereas ascriptions of moral predicates to things that have nothing to do with justice and the common good (hand-clapping, for example) would lack moral content.

There are two problems with this theory. First, it implies some counterintuitive conclusions. For example, it suggests that propositions of the form 'X is morally permissible' would not have moral content whenever X has nothing to do with justice or the common good. The same could be said for such propositions whenever they are believed or expressed by someone who does not believe that ethics has anything to do with the common good or justice. Such a person might, for example, be an ethical egoist

who believes that truly moral action consists in the strong subverting the weak for their own benefit. When this individual takes him or herself to believe that something is morally permissible, for example, he or she does not, according to the view under discussion, believe a moral proposition because this proposition has nothing to do with justice or the common good. According to this view, the ethical egoist is therefore committed to an incoherent theory insofar as ethical egoism is confused about moral concepts. A similar claim is true for the normative moral view that moral facts counterfactually depend upon the attitudes and sentiments that humans have towards certain acts (several such views will be discussed in chapters two and three). These individuals would also claim that justice and the common good have no necessary connection to morality, as it is entirely possible for people to have wildly divergent attitudes about justice and the common good. According to Foot, individuals who endorse a subjectivist theory such as this one are also incapable of coherently making moral judgments. This is problematic, as it suggests the counter-intuitive conclusion that many moral debates are actually the result of conceptual confusion on the part of one or both of the interlocutors. It seems more accurate to say that the egoist and subjectivists considered above are saying something that is false for *substantive* moral reasons, rather than that they are saying something incoherent.

The second problem with Foot's claim is that it relies explicitly on a substantive normative theory. According to this theory, we cannot determine whether a proposition is moral without first having an answer to the questions 'what is just?' and 'what is the common good?' But in order to determine the best way to characterize terms like 'justice' and 'the common good' we must have a prior theory of the application of these concepts.

Since these are both philosophically complex questions, Foot's requirement substantially complicates the task of determining whether or not a specific proposition has moral content.

If it is possible for us to determine the difference between moral and non-moral propositions in a way that is neutral with respect to as many moral questions as possible, then we should pursue this goal. In addition to the fact that a morally-neutral test would simplify the task of determining whether or not a proposition has moral content, it has the added benefits of allowing proponents of different moral theories to potentially agree on such an account and minimizing the number of contentious presuppositions that our theory of moral content relies upon. For these reasons, I will continue to assume that theories like Foot's do not provide us with a satisfactory account of moral content.

I.iii – Introducing My Proposal: MORAL

The claims I have made up to this point are not meant to be controversial, but are rather meant to explain how I will use terms and what methodological assumptions will guide my arguments that particular propositions have moral content. Since I have borrowed many of my assumptions from philosophers who have approached this question, the majority of my direct interlocutors should not find anything I have said to be controversial (especially Karmo 1988, Dreier 2002, and Fantl 2006).

As I mentioned above, most philosophers endorse the following as a sufficient condition for having moral content: a proposition P has moral content if P ascribes a

moral predicate to some act or state of affairs. This proposal picks out all and only the uncontroversial set of moral propositions, and is therefore an uncontroversial sufficient condition for having moral content. Much of the philosophical work on this topic has attempted to expand upon this sufficient condition by developing broader definitions of the distinction between the moral and the non-moral. These definitions are usually in the form of necessary and sufficient conditions. However, my methodology will be to expand upon the one uncontroversial sufficient condition by proposing *only* an additional sufficient condition for which the following two things are true: 1) it is only satisfied by propositions that are not paradigmatic examples of non-moral propositions, and 2) it is satisfied by a number of propositions that the uncontroversial sufficient condition does not identify as moral, but that there is nonetheless reason to think have moral content (e.g., some compound propositions composed of moral and non-moral atomic propositions and identity claims that identify a moral property with a non-moral one). In doing so, this proposal explains why some compound propositions, some property identifications, and other propositions seem to belong in the 'moral' category even though they don't satisfy the uncontroversial condition. My proposal is therefore more modest than that put forward by other philosophers, but it nonetheless will help inform debates about the disputed cases. Specifically, it will inform the arguments I develop in later chapters in favour of the claim that theories about the nature of moral judgment have moral content.

To summarize, the following statement is an accurate characterization of one of my major methodological assumptions: if a test (T) successfully identifies a class of propositions that are thought to be moral but that do not satisfy the uncontroversial

sufficient condition *and* T does not identify paradigmatically non-moral propositions as being moral, then there is good reason to believe that all propositions that satisfy T have moral content. Any proposal that satisfies the conditions set out in the antecedent of this conditional will therefore be useful for the purpose of resolving debates about whether or not any particular propositions have moral content.

I believe that the following proposal satisfies the conditions set out in the antecedent of that conditional:

> MORAL: A proposition P has moral content if there exists some proposition Q that has moral content and believing both P (not one of the entailments of P) and Q lead one to make contradictory moral judgments about a particular situation (provided the description of that situation does not include any moral information)

In other words, MORAL states that if believing a proposition P in a situation commits me to contradicting another possible moral judgment[9] about that situation, P has moral content. By way of illustration, MORAL can also be understood as the claim that any proposition that can add moral information to a situation, (and thus rule out certain moral claims about that situation) without the addition of further moral assumptions, must have moral content.

For example, take any situation that can be described without the inclusion of any moral information. One such example might be a situation where I am deciding whether I should communicate a true proposition (which, if believed by the listener, will create a

[9] In this context (and in MORAL) by 'moral judgment' I simply mean an ascription of a moral predicate to an action or a state of affairs. I do not mean for this stipulation to be taken as a substantive theory about the nature of moral judgment.

surplus of pain in the long term), or a false one (which, if believed by the listener, will create a surplus of pleasure in the long term) to a listener who will believe whatever I say. In other words, I am in a situation where I must decide whether to tell a pleasant lie, or reveal a painful truth. Note that my description of this situation does not contain any moral information. However, if I were in this situation and I believed the following proposition:

> L: It is always morally wrong to lie.

then I would be committed to ascribing a particular moral predicate to this situation. Therefore, the uncontroversial sufficient condition identifies L as having moral content. Furthermore, if I believed both L and the proposition:

> H: If an act maximizes the amount of pleasure in the long term, then it is morally right.

I would be committed to believing that telling the lie would be both morally right and morally wrong. I therefore cannot believe both H and L without making contradictory moral judgments about a situation. According to MORAL, H must therefore have moral content. This is an example of how MORAL adds a plausible sufficient condition that explains why some propositions which do not satisfy the uncontroversial sufficient condition (such as H, for example), have moral content.

It is, however, extremely important to note that the situation mentioned in the test *must* be described without the inclusion of any moral information. If we grant other moral background conditions, some non-moral propositions would satisfy MORAL. For example, if one background condition that is built into the description of a situation is the truth of classical hedonistic utilitarianism, then the proposition 'chocolate ice cream induces more pleasure in Betty than does vanilla' contradicts the moral proposition 'Betty morally ought to eat only vanilla ice cream.' But any theory that identifies claims about the relationship between pleasure and ice cream as being moral can be accused of identifying paradigmatically non-moral propositions as moral, and thus cannot be a sufficient condition for having moral content. This is why I include the proviso that the description of the situation we employ when applying MORAL must not include any moral information about the situation (i.e., it must not include any moral judgments about the situation or include any moral propositions as part of the background conditions). I therefore take MORAL to imply that whenever a proposition can add moral information to a situation that is not otherwise described as having any moral qualities (or, in other words, whenever a proposition gives *moral advice)*, this proposition must have moral content.

This proposal may seem bizarre, overly complicated, or confusing. I concede that these may be apt criticisms. I have selected MORAL, however, because unlike all other proposals I have found in the philosophical literature, it does not identify any paradigmatically non-moral propositions as being moral or, in other words, does not generate false positives. MORAL contains so many specific provisos because they are designed to prevent it from identifying too many propositions as moral. It will not serve

as a definition of morality, as there are some propositions that arguably have moral content that MORAL will not pick out (such as, for example, conjunctions of moral propositions). But if Dworkin is correct that no helpful definition of morality is forthcoming, the best we can hope for is a useful tool that can inform debates about disputed cases by identifying some plausible sufficient conditions for having moral content. I argue that MORAL is such a tool, as it generates the intuitively correct results for many disputed cases, and does not identify any false positives.

As another example how MORAL can serve as such a tool, consider the uncontroversially moral proposition:

> LYING1: Lying is morally permissible.

If I were to believe LYING1 while also believing:

> LYING2: The property of being a lie is identical to the property of being morally impermissible.

there are many situations in which I would be committed to concluding that some things (namely, lies) are both morally permissible and morally impermissible. But the uncontroversial sufficient condition (which states that a proposition P has moral content if

it ascribes a moral predicate to some act or state of affairs) does not identify LYING2 as having moral content. MORAL, however, does identify it as having moral content. There is a clear sense in which LYING2 gives moral advice that conflicts with the uncontroversially moral advice given by LYING1. If I believed both of these propositions, I would be committed to contradictory moral judgments about many situations. Because of this feature of LYING2, we can conclude that it has moral content.[10]

In later chapters, I appeal to MORAL while defending my central thesis that philosophical theories about the nature of moral judgment have moral content. However, before I can turn to these arguments, I must address a number of questions about MORAL. In particular, I need to further explain the mechanics of its application, and about my motivation for defending it instead of other alternatives. I will first survey several other theories that philosophers have proposed about what separates the moral from the non-moral. After arguing that they all succumb to a particular counterexample, I will explain what I take to be the major virtues of MORAL. The arguments in these sections are largely independent of the arguments in later chapters, as those chapters can probably succeed by relying only on the intuitive, common sense conception of the distinction between the moral and the non-moral that Dworkin and others employ. If you are convinced already that MORAL is correct, or you are simply not interested in the details of this account, you could skip the following sections and proceed to Chapter Two.

---

[10] Dworkin similarly argues that such property identifications are just re-statements of moral propositions: "There is no difference in what two people think if one thinks that the only thing that can make an act right is its maximizing power, so that it makes no sense to evaluate rightness in any other way, and the other thinks that the property of rightness and the property of maximizing power are the very same property. The second opinion uses the jargon of metaphysics, but it cannot add any genuine idea to the first, or subtract any from it. It sounds more philosophical but it is no less evaluative" (Dworkin 1996 p. 101).

II –Alternative Theories

A number of philosophers have proposed tests for determining whether a proposition has moral content. For the most part, these philosophers have attempted to defend *definitions* of the moral domain, rather than the kinds of sufficient conditions I described above. In this section, I will show that none of these proposals can be used as a helpful sufficient condition, as they all identify too many propositions as having moral content.

One intuitively plausible and influential test for differentiating moral from non-moral propositions is often attributed to David Hume. The passage that is generally appealed to is as follows:

> For as this *ought*, or *ought not*, expresses some new relation or affirmation, 'tis necessary that it should be observed and explained; and at the same time that a reason should be given; for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. (Hume, *T* pp. 520/III.I.i)

This proposal, which is commonly referred to as Hume's Dictum or Hume's Law, is usually clarified in the following way:

> A proposition has moral content if and only if it ascribes some moral predicate to some act or state of affairs or entails a proposition that ascribes some moral predicate to some act or state of affairs

This claim is equivalent to the thesis that there is no logically valid argument in classical logic with non-moral premises and a moral conclusion. This plausible test is based on the intuitive idea that any proposition that is logically inconsistent with a moral proposition must itself have moral content. This test is still popular and is defended by some contemporary philosophers (see, for example, Kramer 2009 pp. 6-9).

The most commonly noted problem with Hume's Law is that it does not hold for all inferences. A simple argument to this effect was articulated by A.N. Prior (1960). Consider any morally-neutral sentence N and morally-loaded sentence M. When we think about the disjunction 'N v M' we face a dilemma. If 'N v M' has moral content, then Hume's Law is false because this moral disjunction is derivable from the non-moral premise N. However, if 'N v M' does not have moral content, then we can construct a valid argument with the premises 'N v M' and ~N with the conclusion M. I other words, regardless of how we understand mixed disjunctions, we can derive a moral proposition from a non-moral one. Hume's Law therefore fails as a definition of the moral domain.

The most common solution to the dilemma is to relativize the notion of moral content by developing an account of how a proposition can have moral content in some situations but not in others. This line of argument has led some philosophers to reject Hume's Law in favour of a modified criterion. Toomas Karmo, for example, has suggested that the property of expressing a moral proposition is world-relative: sentences might express moral propositions in some possible worlds, but express non-moral

propositions in other worlds (Karmo 1988). According to Karmo's proposal, a proposition has moral content at a world only when we cannot determine its truth value at that world without moral inquiry. As Karmo puts it, "we define a sentence S to be ethical in a possible world w just in case S is true in w with respect to one ethical standard, and false in w with respect to another ethical standard" (Karmo 1988 p. 254).[11] According to Karmo, then, the mixed disjunction 'N v M' has moral content in any world where N is false, but lacks moral content in any world where N is true. This is because whenever N is true, we do not have to do any further moral investigation to know whether 'N v M' is true, since any disjunction with a true disjunct is guaranteed to be true. However, whenever N is false, we cannot know the value of 'N v M' until we engage in moral inquiry and thus endorse at least a partial ethical standard. A proposition is therefore Karmo-moral (as some philosophers discussed below will use this word) in a possible world whenever we cannot determine its truth without additional moral information. In the world where N is true, we can determine that 'N v M' is true without needing to engage in moral inquiry. It is therefore not Karmo-moral in that world.

Karmo takes this view to imply a re-articulated version of Hume's Law according to which there are no *sound* arguments with only non-moral premises and moral conclusions. This means that if we find a set of true premises (in a world) that logically entail a moral proposition, then at least one of the members of this set must have moral content relative to that world (Karmo 1988 p. 256). I will argue below that this element of Karmo's view, along with all the other views here canvassed, suffers from the problem that it identifies too many propositions as having moral content. There are, however,

[11] 'Ethical standards' are, according to Karmo, complete truth-value assignments for the set of uncontroversially moral propositions.

several other problems with Karmo's view. I introduce these problems now because later in this chapter I will show that MORAL does not succumb to the criticisms other philosophers have raised for Karmo's account.

The primary problem that many philosophers have raised for Karmo's thesis is that it identifies all conditionals whose antecedents are false at a particular world as lacking moral content in that world. (Again, this is because any material conditional with a false antecedent is guaranteed to be true given the truth table for the horseshoe.) There is, however, good reason to think that these conditionals can have moral content in those worlds (these mixed conditionals, much like the mixed disjunctions described above, are paradigmatic examples of the 'disputed cases' I described above.) A number of philosophers have pointed out that conditionals with false, non-moral antecedents and moral consequents seem to be giving some kind of moral advice (See, for example, Fantl 2006, Dreier 2002, Alm 2000, Blackburn 1988). This view is supported by considerations like the following. The proposition

> ANIMALS1: It is always morally permissible to eat food products that come from animals.

seems to be giving some kind of moral advice regarding what you are morally permitted to eat. It seems obvious that the proposition:

> ANIMALS2: It is always morally impermissible to eat food products that come from animals.

is in tension with ANIMALS1. They are giving inconsistent moral advice. The conditional proposition:

> ANIMALS3: If that came from an animal, then it is morally impermissible to eat it.

is different in form from both ANIMALS1 and ANIMALS2, but it still seems to be giving moral advice. In fact, it appears to be giving exactly the same moral advice as ANIMALS2. The problem is that according to Karmo's theory of the semantics of conditionals (according to which all conditionals are interpreted as material conditionals),[12] whenever the food item under discussion did not come from an animal, ANIMALS3 is true. In these cases, Karmo is committed to the claim that ANIMALS3 lacks moral content. But that seems implausible. Any satisfying account of moral content should be able to explain why ANIMALS3 seems to have moral content that seems to be in tension with the moral content from ANIMALS1, and Karmo's theory cannot offer any such explanation.

[12] This is implicit in Karmo's work, but is made explicit in Fantl's (2006) discussion of Karmo's and Dreier's theories. Fantl says, for example, that "read as a material conditional, the statement *If your hand is holding a loaded gun aimed at a harmless person, then it is good to move your trigger finger in a shooting motion* is true [when the antecedent is false]" (Fantl 2006 p. 27). This demonstrates that Fantl and others in this tradition are assuming an account of counterfactual conditionals that treats them as material conditionals. It is true that the adoption of a different theory of the semantics of counterfactual conditionals might resolve this problem, but such a revision would alter the point and scope of Karmo's project. In addition to this point, a detailed discussion of the semantics of conditionals would simply take me well beyond the scope of this chapter. My own proposal, MORAL, does not presume a particular theory of semantics of conditionals. I raise this criticism now only to provide the necessary background for later sections of this chapter.

In order to remedy the deficiencies in Karmo's theory, James Dreier has proposed a third test. Rather than relativizing moral content to possible worlds, Dreier relativizes moral content to the beliefs of individuals making assertions. Different tokens of the same type of assertion may have moral content relative to one speaker, but lack moral content relative to another. For example, consider again the mixed disjunction 'N v M.' A token assertion of 'N v M' would *not* have moral content as long as the individual making the assertion believes N. This is because the individual in question can believe 'N v M' in virtue of the fact she believes N. In this case, her belief that the disjunction is true says nothing about her attitude towards M. The same is not true for the individual who believes ~N but nonetheless asserts 'N v M,' as this individual's assertion commits her to the moral claim M. Note the similarity with Karmo's account: both Dreier and Karmo rely on basic facts about the truth-conditions of disjunctions, and reason from the fact that a disjunction of the form 'N v M' could be made true entirely by the non-moral N to the conclusion that whenever 'N v M' is true/believed *because* N is true/believed, then the disjunction (or, for Dreier, the assertion of the disjunction) lacks moral content.

Dreier originally stated his proposal in the following way: "asserting a proposition commits you morally whenever that proposition is Karmo-moral [it satisfies Karmo's definition cited above] relative to the world the speaker believes herself to be in" (Dreier 2002 p. 253). In other words, an assertion has moral content according to Dreier when one cannot assert it without committing oneself, relative to the beliefs one holds about the nature of the world, to other uncontroversially moral propositions. At other times, Dreier formulates his test with an emphasis on the nature of a 'moral commitment:' "what matters to the conception of a 'moral commitment' (of an assertion) is what a person's

view must be in order for the person to be willing to make the assertion" (Dreier 2002 p. 248). According to this statement of his thesis, if the possible moral views that are available to an individual are limited or constrained by the fact that he or she was willing to make a particular assertion, then that assertion must have had moral content. In other words, if an assertion reveals something about a person's moral beliefs, given his or her non-moral beliefs, that assertion had moral content.

Fantl articulates Dreier's view more rigorously with the following definition: "P commits you to Q just in case your willingness to assert P means that you must believe Q in order to remain ideally rational" (Fantl 2006 p. 29) where 'ideally rational' does not mean fully informed, but rather means that you have not made any mistakes in reasoning (e.g., you have not said anything that seems paradoxical, contradicted yourself, or violated first-order logic). According to Fantl's version of Dreier's view, any time an assertion commits the speaker to a moral proposition, the assertion must be understood as having moral content for that speaker. It seems that this is the version of the criterion that Dreier settles on, as he states that any assertion that reveals that I do or do not accept some subset of possible moral beliefs will be morally committal:

> If I assert a statement that is true relative to some moral standards but not others, you will be able to draw some conclusions about my moral standards. I will have committed myself to standing by one or another of those moral standards that count the statement true. So as you listen to me make assertions, you will be able to narrow down the class of moral standards that could make my assertions true. (Dreier 2002 p. 254)

Note that, as Dreier states his view here, any assertion that commits me to affirming or denying some set of moral propositions will be morally committal for me insofar as it forces me to 'stand by' one of the moral standards that affirms or denies the propositions in question. In this way, such assertions 'narrow down' the class of moral propositions that I can believe while remaining ideally rational. It therefore seems plausible to claim that Dreier's view implies the following claim (as stated by Fantl): "a metaethical position fails to be morally neutral if it commits you to *the denial of* some moral statement" (Fantl 2006 p. 29). Whenever anyone commits themselves to denying a moral proposition, we can certainly narrow down the class of moral standards that would make that person's assertions true. If a specific assertion is counted as false by a moral standard, then my willingness to make that assertion demonstrates that I am likewise committed to the falsity of that moral standard. Under Dreier's assumptions, we must assume that any assertion that commits someone in this way has moral content.

Note that this interpretation of Dreier's statement of his view is actually inconsistent with his initial statement of the view: "asserting a proposition commits you morally whenever that proposition is Karmo-moral relative to the world the speaker believes herself to be in" (Dreier 2002 p. 253). According to this original version, if I assert something that commits me to the denial of a moral proposition but is not, relative to my beliefs, true according to one moral standard and false according to another (Karmo-moral relative to my non-moral beliefs), then that assertion does not have moral content. This is inconsistent with Dreier's later claim that any assertion that reveals something about my moral commitments has moral content, as it is possible for a proposition to be false according to one moral standard (and thus rule out that standard as

one that I might endorse), but not true according to any other (i.e., it cannot be derived or otherwise shown to follow from any other). Such a proposition might, for example, just be indeterminate relative to other standards.[13] In such a case, Dreier's two formulations of his criterion seem to give a different verdict.

I am not sure what to make of this inconsistency on Dreier's part, but I will continue to treat the latter statement of his view (the one re-articulated by Fantl) as his considered view. I will do this because this statement of his view seems to more accurately capture his thoughts on the matter, as he frames his discussion in these terms more often and more consistently.

The theories endorsed by Hume, Karmo, Drier, and Fantl are all committed to a claim that opens them up to an objection which I will be referring to as simply, the *reductio*. The commitment in question is that any proposition that is *logically inconsistent* with a moral proposition has moral content.[14] (Two propositions P and Q are logically inconsistent if they cannot both be true at the same time.) This commitment is explicitly central to Hume's Dictum, which directly appeals to the logical entailments of a proposition in order to determine whether or not it has moral content and logical entailment is a concept that can be expressed in terms of logical inconsistency.

Karmo's view carries a similar commitment, although it is somewhat more complicated. It is not the case, according to Karmo, that all propositions that are logically

[13] One example of such a sentence is 'some peas are green', or: ∃x(x is a pea & x is green). I demonstrate below that this sentence is logically inconsistent with a conjunction of moral propositions (and therefore is false according to at least one moral standard) but does not follow from any internally consistent moral standards (and therefore is indeterminate with respect to those moral standards that it does not contradict).
[14] We could also follow Fantl in adding the proviso that this is only true when neither of these propositions are self-contradictory or inconsistent: "Let us exclude these self-contradictory and (for lack of a better term) self-defeating moral statements" (Fantl 2006 p. 25).

inconsistent with a moral proposition will be false according to at least one moral standard and true according to another (see below example). This statement of Karmo's view does not, therefore, imply that all propositions that are logically inconsistent with a moral proposition have moral content. However, Karmo's view also requires that for all sets of sentences, at least one member of that set has moral content in a world whenever the following two conditions obtain in that world: 1) the set entails a proposition that is moral in that world, and 2) all of the members of the set are true in that world.[15] Since we can explain entailment in terms of inconsistency, Karmo is therefore committed to the slightly different claim that any proposition that is *true at a particular world* and is logically inconsistent with an uncontroversially moral proposition has moral content *in that world.* (This is because the proposition under discussion will also entail a moral proposition, namely, the negation of the moral proposition with which it was inconsistent. Such negations have moral content in all possible worlds, as they are always true according to one moral standard and false according to another.) For the purposes of my argument, this difference between Karmo and the others is superficial and will not substantially impact my arguments.

Dreier and Fantl, however, both tacitly endorse the same version of the commitment as Hume. According to Dreier (or at least the version of Dreier's proposal articulated by Fantl), any time my acceptance of a proposition P logically commits me to the denial of a moral proposition Q, this acceptance reveals something about my moral view – namely, that it is a moral view that rejects Q, and this is sufficient for concluding

[15] Here is Karmo's statement of this view: "if sentences Si,…,Sn (Where n > 0) entail sentence S(n + 1), then for any possible world w in which S(n + 1) is ethical, if all of Si,…,Sn are true in w, then at least one of Si,…,Sn is ethical in w" (Karmo 1988 p. 257).

that P has moral content. This means that according to Dreier's theory, a proposition that is logically inconsistent with a moral proposition will have moral content relative to every possible set of beliefs. The views I have here canvassed, then, imply that all propositions that are logically inconsistent with moral propositions have moral content (with the exception of Karmo, who must nonetheless endorse a somewhat modified version of this commitment).

This commitment implies that for any moral proposition (M1), its negation (~M1) will have moral content. This implies that any conjunction containing two moral propositions (M1 & M2) will also have moral content, as the conjunction will be logically inconsistent with the negation of the original proposition (~M1). In other words, since (~M1) has moral content, so does (M1 & M2). And since (M1 & M2) has moral content, so does ~(M1 & M2), as these last two propositions are also logically inconsistent. As Fantl puts it, an obvious consequence of the assumption that any proposition that is logically inconsistent with a moral proposition has moral content is that if a proposition "commits you to the denial of a conjunction of moral statements, it is not morally neutral" (Fantl, 2006 p. 27).[16]

The problem for these views, then, is that they identify too many propositions as having moral content. Our paradigmatic example of a non-moral proposition discussed above was 'some peas are green.' However, all of the views discussed above imply that 'some peas are green' has moral content. According to all three of them, 'it is wrong to eat peas and it is not wrong to eat green things' has moral content. According to Hume's

[16] Note that this argument runs somewhat differently for Karmo. For Karmo, each of M1, ~M1, M2, ~M2, (M1 & M), and ~(M1 & M2) must have moral content in all possible worlds simply because in all possible worlds they are true according to one moral standard and false according to another.

view, this proposition has moral content because it logically entails moral propositions. For Karmo, it has moral content in all possible worlds as it will always be true according to one moral standard and false according to another. Similarly, according to Dreier it has moral content relative to all possible beliefs about the way the world is, as any assertion of this proposition will reveal something about the speaker's moral view. However, this proposition is logically inconsistent with 'some peas are green.' More formally, the propositions:

1) ∀x(x is a pea → it is wrong to eat x)

and

2) ∀x(x is green → it is not the case that it is wrong to eat x)

both have moral content. According to the commitment I ascribed to Hume, Karmo, Dreier, and Fantl, this means that the conjunction of (1) and (2):

3) ∀x(x is a pea → it is wrong to eat x) & ∀x(x is green → it is not the case that it is wrong to eat x)

also has moral content. Furthermore, this conjunction is logically inconsistent with the proposition:

4) $\exists x$(x is a pea & x is green)

Therefore, according to Hume, Karmo, Fantl and Dreier 'some peas are green' ($\exists x$(x is a pea & x is green)) has moral content.[17] The four views canvassed above all imply that some paradigmatically non-moral propositions have moral content. I take this consequence to be a *reductio* of these theories.

I now move on to a discussion of a sufficient condition for a proposition having moral content that does not succumb to this criticism, namely, MORAL. One of the major reasons I have discussed views like Karmo's, Dreier's, and Fantl's, is that their work indicates what kinds of features would make a test for moral content successful in the eyes of interested philosophers. I will argue below that many of the concerns raised by these philosophers (as well as the concerns I just raised) are addressed by MORAL.

[17] For Karmo, it has moral content only in our world and in worlds that are relevantly like ours. This is because an argument with (4) as a premise and the negation of (3) as a conclusion is sound in our world. This means that (4) has moral content in our world (and any other world where at least one green pea exists) because it is the sole member of a set of sentences that entails a proposition that is ethical in this world and it is true in this world. Therefore, according to Karmo, 'some peas are green' is moral in our world, despite the fact that it is a paradigmatic example of a non-moral proposition.

III – Virtues of and Problems for MORAL

It is essential for my project and for many other philosophical debates that we have at least a provisional notion of what sorts of propositions count as having moral content. While there is agreement about a large set of uncontroversially moral propositions, there are still many disputed cases, and many philosophers have offered proposals about how to resolve these disputes. I believe I have shown above that some of the most prominent accounts fail to contribute useful sufficient conditions that are satisfied only by moral propositions. The remaining chapters of this dissertation all make use of the distinction between moral and non-moral propositions in order to show that a number of propositions which are widely thought to be morally neutral must be understood as having moral content. Before moving on, then, it is important that I clarify how I understand MORAL, as I will employ it as a guide for the remaining chapters. I will not always explicitly apply MORAL, but my acceptance of it nonetheless shapes and informs the arguments in later chapters.

The remainder of this chapter is therefore aimed at further clarifying how I understand the mechanics of MORAL, and pointing out why it is sensitive to many of the concerns raised by Hume, Prior, Karmo, Dreier, and Fantl.

First, recall how I framed MORAL:

> MORAL: A proposition P has moral content if there exists some proposition Q that has moral content and believing both P (not one of the entailments of P) and Q would lead one to make contradictory moral judgments about a particular

> situation (provided the description of that situation does not include any moral information)

The idea that is central to MORAL is that having moral content is not a formal property of a proposition but rather has to do with the way we can use the proposition in moral deliberation. By denying that moral content is a formal property I mean to claim that moral content is not a matter of a proposition's logical form or its entailment relations, but is better thought of as a matter of its potential contribution to moral practice. If it is the sort of proposition (not its entailments) that can be used to give moral advice without the addition of further moral premises, then it has moral content.

I now turn to an in-depth defense of MORAL. I will provide three distinct arguments in support of the claim that MORAL successfully identifies a sufficient condition for concluding that a proposition has moral content. First, MORAL gives the intuitively correct results regarding conditionals and other problematic cases. Second, it does not succumb to the *reductio* I raised for the four other views. Third, it explains why some problems for Hume's law seem persuasive and it preserves the central intuition that makes Hume's law seem so plausible.

III.i – Mixed Compounds

The first virtue of MORAL is that it seems to give many of the intuitively correct results regarding which mixed compound propositions have moral content. As noted above, counterfactual conditionals with non-moral antecedents and moral consequents are

often understood as having moral content. One virtue of MORAL is that it is sensitive to this fact. Consider, for example:

> GUN1: If my hand is holding a loaded gun aimed at a harmless person, it is morally permissible to pull the trigger.[18]

I cannot believe this proposition simultaneously with:

> GUN2: It is morally impermissible to shoot a loaded gun at a harmless person.

without ascribing contradictory moral predicates to the same act. Since GUN2 has moral content, so must GUN1. I take this example, in conjunction with the examples I detailed above, to be sufficient evidence that MORAL gets the intuitively correct result for mixed compounds of this sort.

III.ii – The Reductio *and The Proviso*

The second virtue of MORAL is that it cannot be refuted by the *reductio* I raised against other contemporary theories. The reason for this is that MORAL is not committed to the claim that any proposition that is logically inconsistent with a moral proposition also has moral content (or any of the varieties of this claim endorsed by Karmo). It is this

[18] This is a modified version of an example found in Fantl (2006).

desirable feature of MORAL that motivates the inclusion of the proviso 'not one of the entailments of P.' I will first explain how I understand this proviso, and then demonstrate how it allows MORAL to avoid the *reductio* I introduced above.

Many theories accept some form of the claim that if any of the entailments of a proposition have moral content, this is sufficient evidence to conclude that the original proposition also has moral content. It is this commitment that the proviso denies. We could rephrase the proviso as stating that if there exists some proposition Q that has moral content and believing both Q and *one of the logical entailments of* P would lead one to assign contradictory moral predicates to an act or state of affairs, then this information is *not sufficient* for determining that P has moral content. If a proposition R is entailed by a proposition P, and MORAL identifies R as having moral content, this would not be sufficient evidence to conclude that P has moral content. This means that MORAL does not pick out conjunctions that contain moral propositions as having moral content.[19]

This resolves some of the problems that I raised for the other views, insofar as MORAL does not treat conjunctions of moral propositions as having moral content. The fact that paradigmatically non-moral propositions are sometimes logically inconsistent with conjunctions of moral propositions does not, according to MORAL, imply that those

[19] Note that the case of conjunctions is very different from the mixed disjunctions and conditionals I discussed above. In the case of a conditional, the antecedent describes a situation and the consequent renders a moral verdict about this situation. A similar characterization apples to mixed disjunctions, as one disjunct describes a situation and the other states that a particular moral proposition must be true whenever that first disjunct is false. In these cases, it is the *entire compound proposition* that contributes moral information, not an entailment of the proposition. In the case of conjunctions (either mixed conjunctions, or conjunctions of nothing but moral propositions) the atomic propositions do not combine to offer any moral advice that goes beyond the moral advice given by the individual conjuncts. The conjunction itself does not contribute any moral information, but the individual conjuncts do. The proviso therefore does not exclude mixed conditionals and disjunctions, as it is not the entailments of those propositions that add moral information, but the entire compound propositions (even though, in the course of using those mixed conditionals/disjunctions to generate moral advice, you will have to perform some logical operations on them and thus derive their entailments).

paradigmatically non-moral propositions have moral content. However, this feature of MORAL raises another problem, namely, that the proviso is simply counterintuitive, as it seems obvious that many conjunctions of moral propositions do have moral content. Like many of the other theories above, it seems possible that I have misdrawn the boundaries of the moral domain. However, I believe there are three plausible responses to such worries.

My first response to this worry is that I am not committed to denying that *all* conjunctions of moral propositions lack moral content. Some such conjunctions might actually be elliptical for a moral proposition. For example, consider:

> JIM/HARRY: It would be morally wrong for Jim to steal that car and the same is true for Harry.

It seems reasonable to think that this conjunction, while it consists of two propositions, seems to be giving moral advice and ascribing a moral predicate to only one action: the stealing of the car.[20] If we understand JIM/HARRY this way, MORAL would identify it as having moral content. While I think this is plausible, it seems somewhat ad hoc to conclude that this conjunction is ascribing only a single moral predicate. It is equally plausible that it is ascribing two predicates: one to the act of Jim stealing the car and another to the act of Harry stealing it.

[20] I am counting the two token actions of Jim stealing the car and Harry stealing the car as being, for these purposes, one action, namely, the stealing of a particular car.

Second, it is not clear to me that *all* conjunctions of moral propositions intuitively seem to have moral content. For example, consider contradictory conjunctions like the following:

CONTRADICTION: Abortion is morally wrong and abortion is not morally wrong.

This is an example of a conjunction of two propositions that, in isolation, would certainly have moral content. However, it is reasonable to doubt that the whole conjunction does, as it does not seem to offer any concrete advice or add any relevant moral information. CONTRADICTION does not seem to settle any moral questions, and so it is plausible to think of it as lacking moral content. Cases like this demonstrate why we may not want a criterion like MORAL to pick out all conjunctions of moral propositions as having moral content.

A third possible response is that there may be some additional criterion that gives us reason to conclude that conjunctions of moral propositions have moral content. Recall that MORAL is only meant to be one of many possible conditions that are sufficient for determining that a proposition has moral content. I did not claim that it is the *only* such sufficient condition. The fact that MORAL does not pick out a particular proposition as moral cannot count as evidence against it any more than a similar argument would count against the uncontroversial sufficient condition I described above. MORAL would only be disproved if it could be shown to identify an obviously non-moral proposition as being

moral. This is how I demonstrated that the theories discussed above could not be sufficient conditions for having moral content. Those theories generate false positives, whereas MORAL generates only false negatives.

To summarize, the *reductio* I leveled against other theories does not apply to MORAL. This is in part because MORAL contains the proviso regarding entailment, and in part because it is only meant to be understood as a sufficient condition for having moral content, rather than a definition of the moral domain.

III.iii – Converse Entailment and Moral Advice

The third virtue of MORAL is that it serves to explain why some other criticisms are efficacious against alternative theories. Take, for example, Prior's original complaint against Hume's Law (mixed disjunctions are problematic for Hume's Law, because regardless of whether or not we understand them as having moral content, they serve as a counterexample to Hume's Law.) Prior's dilemma is predicated on the claim that any proposition that logically entails a moral proposition has moral content. It is no surprise that this claim leads to a dilemma, as it has been shown to be mistaken. If we were to commit ourselves to the premise that Prior relies upon, we will wind up having to concede that virtually all propositions have moral content. Once we accept that there is a meaningful distinction between the class of moral propositions and propositions that entail moral propositions, Prior's dilemma ceases to be a dilemma. The mixed disjunction in question simply has moral content, regardless of its entailment relations. For example, if I believe the following two propositions:

ROB1: Either I am not thirsty or it is morally permissible to rob the liquor store.

ROB2: It is always morally impermissible to rob the liquor store.

there will be certain situations in which these propositions would give conflicting moral advice. When I am thirsty, ROB1 will tell me that it is permissible to rob the liquor store, whereas ROB2 will tell me that it is impermissible to perform that same action in identical circumstances. The fact that ROB1 is entailed by a non-moral proposition is irrelevant to determining whether or not it has moral content. Instead, as I have argued, the only relevant factor is whether or not a proposition satisfies MORAL.

This conclusion is, in effect, a rejection of the popular and convincing idea that the class of moral propositions is closed under converse entailment. The most appealing feature of the claim that the class of moral propositions is closed under converse entailment is that it explains why one cannot derive moral advice from non-moral propositions alone. But this is not a mark against my account, as it can also explain why non-moral propositions cannot be used to give moral advice.

The following explanation demonstrates how MORAL is sensitive to this kind of concern. If I believe the non-moral disjunct of a mixed disjunction, then my belief in that disjunct commits me to an infinite number of disjunctions, including all combinations of the original disjunct and a moral proposition. According to MORAL, these disjunctions must all be understood as moral propositions. The obvious problem with this is that I, in

virtue of having certain beliefs, am committed to a nearly infinite number of propositions with moral content, many of which conflict with one another. We can resolve this problem by noting that beliefs acquired in this way cannot in practice offer moral guidance to the person who acquired them. If I accepted 'N v M ' because I believed the non-moral proposition 'N,' then I effectively endorse the moral standard 'whenever N is not true, M is.' However, I am only committed to this moral standard because I believe that 'N' is in fact true. If I am correct in the belief that 'N' is true, then the disjunctive belief 'N v M' will never, in practice, commit me to the claim that 'M' is true. If it turns out that I was wrong and I find out that 'N' is false, then I would no longer be committed to 'N v M' because I would have rejected the premise that initially led me to accept the moral standard.

In short, the only moral advice that can be logically derived from non-moral sources is advice that could never be relevantly applicable, since this advice would only apply when the belief that grounded it turned out to be false. My view can therefore explain and vindicate the most attractive feature of Hume's Law while conceding that many non-moral propositions logically entail moral propositions.

IV- Conclusion

It is clear that we are in need of an account of the difference between moral and non-moral propositions. A number of philosophical debates assume a meaningful distinction between these two classes of propositions, but relatively few have attempted to defend a particular account of what differentiates them. Some philosophers have recognized the importance of this question, but all of the proposals to date suffer from the problem that they identify paradigmatically non-moral propositions as having moral content. I propose that we stop thinking of moral content as being determined by a proposition's entailment relations. Instead, we should focus on whether or not we can understand the proposition as offering moral advice or adding moral information. MORAL captures this claim in the form of a sufficient condition for having moral content. Any proposition that can be shown to satisfy MORAL, I contend, must be understood as having moral content. This proposal can avoid the *reductio* that undermines other contemporary proposals, can explain the role that disjunctive and conditional propositions play in moral discourse and can vindicate the central intuition that buttresses the claim that moral discourse is closed under converse entailment. It is therefore a plausible candidate for being a sufficient condition for having moral content. Since no useful definition of the distinction between the moral and the non-moral seems to be forthcoming, this may be the best we can hope for.

In the second chapter, I offer my first arguments that a number of philosophical theories about the nature and content of moral judgments must be understood as having

moral content. I then turn in chapter three to the implications this has for attempts to answer philosophical questions about the nature of moral judgment by appealing to empirical, scientific evidence.

Chapter Two

Theories of Moral Judgment as Normative Moral Theories

In the last chapter, I argued that the best way to resolve disputes about whether a disputed proposition has moral content is to determine whether or not the proposition can be used to give moral advice or contribute moral information in any situations. In this chapter, I employ these conclusions in order to show that many theories about the psychological nature and content of moral judgments are propositions with moral content. These theories play crucial roles as the foundations of metaethical theories and many philosophers erroneously treat them as morally neutral theses. It seems intuitively plausible that these analyses could be understood as anthropological claims about how particular communities engage in moral judgment, but I argue that this thought is mistaken. While it is true that certain anthropological claims about the types of moral system used by a particular culture are morally neutral, theories about the psychological nature and content of moral judgment *simpliciter* almost always carry a moral commitment. The anthropological claims are not, then, equivalent to the analyses used by many philosophers.

I argue that four prominent metaethical theories about the nature of moral judgment must be understood as propositions with moral content. I focus first on Mackie's analysis of the psychological nature of moral judgment. I prioritize my discussion of Mackie over other philosophers because he is one of the most forceful proponents of the claim that theories about the nature of moral judgment can be

understood as second-order, morally neutral starting points. I then turn to a discussion of a form of relativism, ideal observer theory, and neo-sentimentalism. I show that all four of these analyses have moral content despite the fact that the proponents of the analyses believe them to be morally neutral. I conclude with the claim that theories about the psychological nature and content of moral judgments have moral content and address several objections that one might raise against this thesis.

I – Mackie and the Claim to Objectivity

Metaethical work in the twentieth century is generally understood as being focused on morally neutral questions about ethics, rather than first-order ethical questions about what things are good, bad, right or wrong. For example, J.L. Mackie characterizes questions about the merits or faults of a particular moral system as being first-order, normative questions. When characterizing his own methodology, he states that

> What I am discussing is a second-order view, a view about the status of moral values and the nature of moral valuing, about where and how they fit into the world. (Mackie 1977 p. 16)

When discussing the distinction between the two kinds of questions, he writes that "the first and second order views are not merely distinct but completely independent: one could be a second order moral sceptic without being a first-order one, or again the other way around" (ibid). Mackie's second-order project largely consists of a characterization of the nature and content of moral judgments.

His denial of the objectivity of moral values is not to be conflated with speaker subjectivism. The latter is a view about what moral terms mean, namely, that they are reports of the speaker's attitudes. This theory holds that moral sentences such as 'murder is morally wrong' can be translated, without loss of meaning, into a description of the attitudes the speaker holds towards murder.

Mackie's subjectivism, on the other hand, is a negative ontological claim. Mackie means only to establish the negation of the claim that there "exist relations of a certain kind, objective values or requirements, which many people have believed to exist" (Mackie 1977 p. 17). The claim to objectivity that Mackie is discussing is most lucidly explained as the belief that moral values are part of the fabric of the world, or that values akin to Plato's Forms exist in the same way as trees and human beings.

The claim that moral judgment presupposes the objectivity of moral values is Mackie's proposed analysis of the psychological nature of moral judgments. Mackie begins his discussion by looking at the psychological nature of moral judgment, and then draws conclusions about what kinds of things would have to exist for the moral properties that are the objects of moral judgments to exist:

> I conclude, then, that ordinary moral judgments include a claim to objectivity, an assumption that there are objective values in just the sense in which I am too concerned to deny this. (Mackie 1977 p. 35)

Mackie's central claim is that the following commitment is built in to moral discourse:

MAC: Moral properties are objective properties.

MAC is a consequence of a psychological analysis of the commitments of ordinary moral discourse. According to MAC, the only properties the instantiation of which would make moral propositions true are objective properties. Moral properties, if any exist, have the characteristic of being objective properties.

It is important to note that there are several possible readings of MAC. One reading understands MAC as implying an existential claim that moral properties exist and that they are objective. There is, however, an equally intuitive reading of MAC according to which it carries no such existential commitment. Just as I can claim that unicorns are single-horned horses without implying the existence of unicorns, I am able to claim that moral properties are objective without implying the existence of moral properties. Furthermore, since Mackie aims to explicitly deny that objective moral properties exist, we must presume that he understands MAC in this latter way.

If we accept this reading of MAC, and Mackie suggests that we must accept it in order to count as individuals who have mastered the use of moral concepts, then we are implicitly accepting that only certain kinds of things can possibly count as moral properties. This implies that MAC commits us to ruling out certain normative moral

theories.[21] One example of such a theory is one that Mackie explicitly rejects, speaker subjectivism. Mackie characterizes this view as claiming that “moral judgments are equivalent to reports of the speaker’s own feelings or attitudes” (Mackie 1977 p. 17). The most common form, and the one that Mackie appeals to in order to exemplify this view, is the claim that ‘this action is right’ just means ‘I approve of this action.’

One relevant feature of this statement of subjectivism is that it is not merely a linguistic or semantic claim but also a substantive moral theory or, in other words, a proposition with moral content. Consider again the property-identification:

> LYING2: The property of being a lie is identical to the property of being morally impermissible.

If I endorse speaker subjectivism, then whenever it is true that I approve of lies, I cannot accept LYING2 without being committed to the claim that lies are both permissible and impermissible. I demonstrated in the previous chapter that LYING2 has moral content, and this example shows that believing both speaker subjectivism and LYING2 would lead me to make contradictory moral judgments in some situations. According to MORAL, speaker subjectivism must also have moral content.

Additionally, I cannot believe speaker subjectivism while also believing MAC. Consider the situation where I believe that a particular action is wrong, but I also find

[21] I understand normative moral theories to be propositions with moral content, specifically universally quantified conditional propositions. I therefore understand normative moral theories in roughly the same way as Shafer-Landau understands moral standards, principles, rules, or laws (Shafer Landau 2003. p. 15).

myself approving of the action. It might not be common for individuals to be in this psychological state, but it is surely possible. For example, an individual who was raised within a racist culture might have such an experience. Upon learning the error of her ways, she would come to endorse an objectivist moral theory that understood racist acts as having the objective property of moral wrongness. When presented with a racist act, this individual would know that according to her objectivist moral theory this act is not morally permissible, but due to her conditioning, she would also find herself overcome with feelings of approval.

My argument that MAC has moral content follows from the following observation: a belief in MAC would commit this individual to ruling out certain possible normative moral theories, namely, speaker subjectivism, because speaker subjectivism denies that moral properties are objective. Because of this, MAC commits her to denying some possible moral judgments about the situation. Specifically, if she endorses MAC, she must acknowledge that the instantiation of subjective properties (e.g., her approval) is never sufficient to make a moral proposition true. She is therefore committed to the claim that facts about her attitudes and sentiments are never sufficient evidence to determine that a particular act has a certain moral property. Whereas speaker subjectivism implies that the racist act in the described situation is in fact morally permissible, MAC denies that this conclusion follows from any of the available facts. MAC therefore contradicts the moral advice given by speaker subjectivism in this situation. I take this to be sufficient evidence for the claim that Mackie's thesis that moral properties are objective properties must have moral content.

Note that this conclusion is distinct from Dworkin's claim that Mackie's error-theory (the conclusion that all moral propositions are false) amounts to a moral proposition (Dworkin 1996). I am not arguing that Mackie's error-theoretic conclusion is moral in nature (although it might well be), but rather that one of his apparently 'second-order' presuppositions about the nature of moral judgment precludes certain first-order moral theories, and thus has moral content.

In the following section, I show that Mackie is not alone in mistakenly believing that a theory about the nature of moral judgments is morally neutral, when the theory in question actually has moral content. I discuss three other prominent metaethical theories and show that despite the fact that their proponents believe them to be morally neutral, they must be understood as propositions with moral content.

II- Relativism, Ideal Observer Theory, and Neo-Sentimentalism

The goal of the above discussion of Mackie was to draw attention to how metaethical theses about the nature and content of moral judgments can have moral content insofar as they can be used to give moral advice. Mackie's theory is only one example of this. Many prominent metaethicists fail to take notice of the fact that their preferred theory about the nature of moral judgments has moral content. In this section I will elaborate on this claim and provide evidence in favour of it.

For example, moral relativism is often treated as a morally neutral thesis about the logic of moral terms. The most well-known proponent of this metaethical approach is

Gilbert Harman. In response to the claim that relativism is, amongst other things, an immoral doctrine, Harman claims that the version of moral relativism he defends “is a soberly logical thesis – a thesis about logical form if you like” (Harman 1975 p. 3).

Harman offers a logical analysis of what the ‘ought’ predicate means in moral discourse or, in other words, what the content of a moral ‘ought’ judgment is. It is important to note that Harman is not discussing all moral judgments, but rather a subclass of what he refers to as internal moral judgments: the class of judgments that “say that someone should or ought to have done something or that someone was right or wrong to have done something” (Harman 1977 p. 5). This class of moral judgments is to be differentiated from the class of judgments about what would be better on the whole, or about whether or not an individual is evil. We can only make inner moral judgments about a person if we suppose that she is capable of being motivated by the relevant moral considerations.

Clearly, inner moral judgments do not exhaust the class of possible moral judgments. However, much of our moral practice can be explicated in terms of discussions about what people morally ought or ought not to do. Even if some dimensions of moral discourse remain non-relative under Harman’s theory, the majority of our moral judgments would be relativistic.

Harman’s argument turns on the claim that whenever an individual makes an internal ‘ought’ judgment, that individual is employing the following four-place predicate:

Ought(A,D,C,M)

where A is an agent, D is a type of act, C is a set of background considerations and M is a motivating attitude. 'Ought(A,D,C,M)' is best read as the claim that "given that A has motivating attitudes M and given C, D is the course of action for A that is supported by the best reasons" (Harman 1975 p. 11).

This is another case where a supposedly morally neutral analysis of the content of a moral judgment can be demonstrated to have moral content. Consider, for example, the following situation: a particular action is an instance of lying and a speaker and listener are debating the moral status of the action. Further, the speaker and listener both have positive attitudes towards lying given their current background conditions, and they are both aware of this fact. If these things are true of these individuals, and the speaker believes Harman's analysis of the nature of moral judgment, then the speaker is committed to advising the listener that she morally ought to lie. In this exact same situation, however, if the speaker believed the following principle:

LYING3: It is never the case that one morally ought to lie.

she must advise that the listener ought not to lie. This statement of relativism therefore contributes moral advice and information because it contradicts the moral advice and

information that a number of other moral propositions would contribute in the same situation. Specifically, relativism contributes the moral information that an individual's attitudes are uniquely relevant to determining what she ought to do. In the situation thus described, this information amounts to moral advice that contradicts the advice given by LYING3. Harman's relativistic analysis of the content of moral judgments must therefore be understood as more than a soberly logical thesis. It must be understood as a proposition that has moral content.

The above two metaethical theories (relativism and Mackie's claim to objectivty) are generally associated with forms of moral anti-realism. However, some versions of moral realism rely on similar theories about the psychological nature and content of moral judgments. One example is the theory put forward by Michael Smith. Smith also endorses the distinction between metaethical and normative questions. He writes that:

> In meta-ethics we are concerned not with questions which are the province of normative ethics like 'should I give to famine relief?' or 'should I return the wallet I found in the street?', but rather with questions *about* questions like these. (Smith 1994 p. 2)

Smith goes on to provide many examples of metaethical questions:

> What does the 'should' in such questions mean? Does it signal that these questions are about some matter of fact? If so, then how do we justify giving one answer rather than another? In other words, what sort of fact is a moral fact? (Smith 1994 p. 2)

He then goes on to claim that he is primarily concerned with addressing metaethical questions (Smith 1994 p. 3). However, the theory of moral judgment Smith defends turns out to itself be a proposition that has moral content.

Smith's claims about the nature of moral judgments rely upon his analysis of the nature of evaluative judgments more generally. Smith's view is that "an evaluative belief is simply a belief about what would be desired if we were fully rational" (Smith 1984 p. 160) and is thus a form of ideal observer theory. This account is then applied to moral judgments by interpreting moral judgments as a special case of evaluative judgments, specifically those judgments with a particular type of content. Smith identifies the relevant content by appealing to certain platitudes about the unique subject matter of morality. The examples he provides of this unique subject matter are: 'right acts are often concerned to promote or sustain or contribute in some way to human flourishing', 'right acts are in some way expressive of equal concern and respect' and the like (Smith 1994 p. 184).

Collectively, Smith's considerations amount to a proposed analysis of the content of a moral judgment. According to this analysis, when we judge that a certain act is morally right, we are committing ourselves to the following two claims: 1) the outcome of the act in question is what we would desire if we were fully rationally, and 2) the criteria identified by a set of moral platitudes are satisfied by the act in question. This analysis can be articulated in the following manner:

> SMI: An act is morally right if and only if the result of the action would be desired if we were fully rational and it is the sort of action that will promote, sustain, or contribute in some way to human flourishing, or is expressive of equal concern and respect.

Smith's account of the nature and content of moral judgment entails SMI. By analyzing the psychological nature of evaluative beliefs and applying this analysis to the special case of moral judgment, Smith commits himself to something that looks very much like a standard that we can use to determine moral rightness.

Consider the moral proposition 'lying is morally permissible.' Whenever I am in a situation where a fully rational version of myself would desire that I not lie and the lie in question would inhibit human flourishing, be inegalitarian and would not be respectful of others, I could not believe both SMI and the moral standard that states that lying is permissible. In the situation thus described, I would have to believe that telling that lie is both permissible and impermissible. Smith's theory about the nature and content of moral judgment therefore contradicts other possible judgments about a variety of situations. As I have argued, this is sufficient for concluding that it has moral content.

Much like Harman and Mackie, Smith's goal is to differentiate metaethical and first-order moral investigations. Based on the text quoted above, it is plausible to attribute to him the belief that his theory about the content of moral judgments (SMI) is morally neutral. However, there is a clear sense in which SMI is a proposition that has moral content. It seems that once again, the distinction philosophers presuppose between first-order moral propositions and theories about the nature of moral judgment breaks down under scrutiny.

There are, however, some possible questions about SMI that must be addressed. One concern that might be raised is that first-order moral theories tend to specify the naturalistic conditions under which a moral property obtains. Examples of moral standards I discussed in the last chapter all take this form. What is novel about SMI is that the conditions it specifies contain a normative term, namely the term 'fully rational.' This differentiates SMI from Harman's and Mackie's analyses, as those analyses contain only naturalistic, non-normative terms. Additionally, it is not plausible that disputes about what constitutes full rationality are resolvable by defining the term 'rational' in naturalistic terms. Smith offers a "summary style" account of rationality similar to that put forward by Williams (1981), but this analysis is not meant to be reductive to purely naturalistic terms.[22]

However, the fact that SMI contains a normative term does not disqualify it from being understood as a normative moral standard. Despite the fact that the questions about rationality seem to be normative, it is possible that considerations about rationality are not relevant to the resolution of moral debates. One reason for thinking that questions about rationality and questions about morality can come apart is that it seems entirely coherent to imagine two individuals who agree on a use of the term 'rational', agree that a particular decision would be the most rational thing to do, but disagree about whether or

[22] A summary-style account is one that is meant only to capture a number of specific platitudes about a particular concept. The concept in question here is practical rationality. Smith endorses the summary of rationality put forward by Williams (1981), with some modifications. Smith's proposal is as follows: "in order to be fully rational an agent must satisfy the following three conditions: (i) the agent must have no false beliefs, (ii) the agent must have all relevant true beliefs, (iii) the agent must deliberate correctly" (Smith 1994 p. 156). Smith's account differs from Williams' primarily in the fact that Smith understands (iii) as requiring rational agents to try to find out whether or not their desires are systematically justifiable. Williams endorsed no such requirement. Smith characterizes this process regarding a particular desire as the attempt to "integrate the object of that desire into a more coherent and unified desiderative profile and evaluative outlook" (Smith 1994 p. 159). He characterizes this process as being similar to Rawls's reflective equilibrium. He then goes on to conclude that "the analysis on offer of a normative reason is, and will forever remain, a non-reductive, summary style analysis" (Smith 1994 p. 164).

not performing the action would be morally right. For example, they might both endorse a maximizing theory of rationality according to which the rational action is that which maximizes expected utility for some individual or group. Let us suppose that these hypothetical individuals also agree that taking a particular course of action, P, would maximize expected utility for the individual or group in question. However, they could still coherently disagree about whether or not taking the course of action P is morally permissible. One of them endorses the view that it is always morally permissible to do what is rational, whereas the other endorses a view that implies that rationality and morality sometimes come apart (universal hedonism would, under these suppositions, be such a view). Despite their agreement about the proper analysis and application of the term 'rational', these two individuals still seem to be engaged in a genuine moral disagreement. The predicate 'is rational' seems to play the same role in moral discourse as any naturalistic predicate insofar as deliberators can coherently disagree regarding the moral relevance of rationality.

Additionally, there are many other normative moral standards that unquestionably have moral content even though they contain a normative term. One example is ideal utilitarianism. This theory identifies moral rightness with the property of being that which maximizes the amount of good in the universe (See Ross 1930 p. 16 and Pickard-Cambridge 1932 p. 72). Despite the fact that this theory is fairly vague (the answer to any moral question depends on what we take to be good) and contains a normative term, it does identify certain kinds of considerations as being relevant to moral thought and identifies others as being irrelevant to moral thought. For example, it is consequentialist

in the sense that it identifies the consequences of acts as being uniquely morally salient[23] and is axiological in the sense that it considers the production of things of value to be a central feature of moral actions. It is therefore inconsistent with non-consequentialist and non-axiological moral theories such as Ross's deontological intuitionism. The fact that ideal utilitarianism contains a normative term does not disqualify it from having moral content.

Another example of a moral standard that contains a normative term is T.M. Scanlon's contractualism. Scanlon holds that "an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced general agreement" (Scanlon 1998 p. 153). Despite the fact that this thesis contains the normative term 'reasonably' it nonetheless "is an idea with moral content" (Scanlon 1998 p. 194). The same is true for moral standards that contain the term 'rational.'

There does not seem to be any reason, then, to grant that moral standards that contain normative terms such as 'rationality' do not have moral content. They seem to be used in the same way as other moral standards in order to give moral advice and add moral information in certain situations, and this is the feature that I identified in the previous chapter as being relevant to determining whether or not a proposition has moral content.

[23] It is worth noting that not all versions of ideal utilitarianism are consequentialist. See for example Johnson (1953). For the purposes of this paper, I will focus on ideal utilitarianism construed as a consequentialist theory.

There is another theory about the nature and content of moral judgments that has gained prominence in recent years. This view has been dubbed *neo-sentimentalism* and is discussed most prominently by Allan Gibbard. The view in question is that:

> GIB: What a person does is morally wrong if and only if it is rational for him to feel guilty for doing it, and for others to resent him for doing it. (Gibbard 1990, p. 42)

Gibbard draws a distinction between broad and narrow senses of morality. Broadly, moral judgments are just judgments about what to do, and thus moral reasoning and practical reasoning are the same thing. Theories of morality understood narrowly are theories about a particular way of going about practical reasoning. Narrowly, moral questions are a subset of the questions of practical reason that focus on things like the reasons we have to perform certain actions, develop certain dispositions, and feel certain emotions. Specifically, narrowly moral theories aim to accurately characterize this subset. It is important to note at this point in the discussion that all of the views discussed above are directed at the narrow conception of morality, not the broad. Furthermore, neo-sentimentalism is meant to be a narrowly moral theory (Gibbard 1990 p. 41). Gibbard's analysis is therefore offered in the same spirit as the others I discussed above.

Another feature that Gibbard's neo-sentimentalism has in common with the other theories mentioned above is that it also has moral content. Consider again the moral proposition:

LYING1: Lying is morally permissible

In a situation where it would be rational to feel guilty over telling a particular lie, I cannot believe GIB and the moral proposition LYING1 without committing myself to the inconsistent moral propositions 'it would be morally impermissible to tell this lie' and 'it would be morally permissible to tell this lie.' Gibbard's analysis, like all of those discussed above, has moral content.

Once again, one might raise the objection that the conditions Gibbard identifies as being relevant to determining whether or not an act is wrong are not naturalistic and thus the analysis does not have moral content. The same considerations I brought to bear on this objection in regards to SMI can be similarly applied to GIB. The fact that a moral standard identifies a normative property, such as being rational to want or feel, does not entail that the standard in question lacks moral content.

I have argued that there is a persistent problem within the methodology of metaethics. Mackie, Harman, Smith and Gibbard have offered theories of the psychological nature and content of moral judgments and all four of these theories are thought by their proponents to be morally neutral. But in all four cases, we can find scenarios in which the analyses can be used to give moral advice. I argue that this is sufficient for determining that these analyses have moral content. In the following section of this chapter, I address several possible objections to this thesis.

III - Objections

Because many philosophers have thought of theories about the nature of moral judgment as morally neutral hypotheses, it will be useful to survey some of the possible objections to my thesis that these theories have moral content. I will address three such objections. First, I will address the argument that the theories discussed above should be understood as anthropological or sociological hypotheses. Second, I will respond to the objection that the theories must be morally neutral because they are consistent with diverse normative theories. Finally, I will address the objection that the theories discussed above do not have implications for action whereas moral propositions do have implications for action. As part of this latter discussion, I will touch on the implications of several forms of internalism for my view.

III.i - Anthropology or Morality?

The thesis I have been trying to establish is that it proves to be very difficult to say much about the content and nature of moral judgments without inadvertently becoming committed to moral theses. This is surprising, as many of the theories discussed above do not intuitively seem to have moral content. On first glance, these theories seem to be anthropological or sociological accounts of the moral systems used by a particular culture, namely, the culture in which the philosophers live. If they are all

best understood as anthropological claims, they should all be equivalent to some empirical claim like the following:

> ANTHRO: Culture A thinks about what to do by making a particular kind of judgment, and these judgments have features X, Y and Z.

I agree that ANTHRO lacks moral content. Regardless of how I fill out the variables A, X, Y, and Z, I do not rule out any possible moral advice or contradict any moral theories. Even if I fill in the variables such that I am characterizing my own culture's method of judging what to do, I do not commit myself to any moral claims insofar as I say nothing about the status of my own society's moral system. However, when philosophers try to express this sort of anthropological hypothesis in the following form:

> ANTHRO*: Moral judgments have features X, Y and Z.

these philosophers very often commit themselves to moral propositions. This equation of anthropological claims with analyses of the nature of moral judgment is often illegitimate for the reasons discussed above. Despite the fact that it is quite easy to make anthropological claims that are morally neutral, it proves to be extremely difficult to analyze the nature of moral judgment *simpliciter* without committing oneself to morally substantive theses.

All four of the theories discussed above attempt to provide analyses of moral judgment *simpliciter*, and these types of analyses are not equivalent to anthropological theses. Analyses with the structure of ANTHRO* have moral content, whereas those with the form of ANTHRO do not.

III.ii - Consistency with Normative Theories

A second reason why philosophers might be tempted to claim that these theories lack moral content is that they cannot be used as moral standards to informatively answer all first-order moral questions. Neither Gibbard's nor Smith's analyses can be applied in order to draw moral conclusions absent an account of rationality, and Harman and Mackie's theories do not seem to preclude any of the most popular substantive moral theories. All of these theories are, for example, consistent with versions of utilitarianism, deontology, contractualism, contractarianism and virtue ethics. This fact is probably one of the major reasons that philosophers have been tempted to think of theories of the nature of moral judgment as lacking moral content. However, this feature of the theories is not sufficient to show that they lack moral content. For example, the proposition:

A: Abortion is morally permissible.

is plausibly consistent with all five of these normative theories, but this fact alone does not imply that A lacks moral content. Rather, A is a paradigmatic example of a moral proposition.

The five normative theories mentioned above are less specific than A. They do not, independently of other information, clearly determine the moral status of abortion. It is in this sense that I say they are less specific. However, this feature of these theories is not sufficient for determining that they lack moral content, as they are all paradigmatic examples of moral theories. Similarly, the four psychological theories discussed in sections I and II are less specific than the normative theories insofar as they do not, absent additional information, determine the status of those theories. The fact that they are less specific, however, does not disqualify them from having moral content.

III.iii - Implications for Action?

A third reason one might think these four theories lack moral content is that it seems to be incoherent to deny that paradigmatically moral propositions such as A have implications for action, whereas it seems coherent to deny that the four theories discussed above have such implications. This suggests that there is some kind of relevant distinction between these two classes of propositions. This argument relies on the intuition that it would be very peculiar for someone to endorse a first-order moral proposition like A and yet deny that this belief commits them to acting in any particular way. For example, imagine that there is an individual who believes the proposition 'eating meat is morally wrong,' but continues to eat meat and professes a further belief

that there are no reasons to stop eating meat. This individual's behaviour would cause most observers a substantial amount of bafflement and may even lead us to think that this individual does not fully grasp the concept 'morally wrong.'

The argument under consideration relies on the claim that people who deny that the four theories of moral judgment discussed above have implications for action have not made any analogous mistake. Unlike in the case of paradigmatically moral propositions such as A, it is not clear that the theories about the nature of moral judgment discussed above have implications for action. Two individuals who believe SMI and GIB respectively may live their lives identically and not disagree about what acts they have most reason to perform without any inconsistency on either of their parts. This in turn suggests that the four theories discussed above do not have moral content because, unlike paradigmatically moral propositions like A, it is not a conceptual truth that they have implications for action.

By the phrase 'implications for action', I mean that propositions like A seem to give us reasons to behave in a certain way. This objection capitalizes on the fact that the four theories discussed above do not seem to give us any such reasons. This objection is therefore committed to the claim that whether or not a proposition has moral content depends entirely on whether or not it has such implications. Under these presuppositions, an individual who accepts A but also believes that this acceptance has no implications for how she should live her life is committed to an incoherent set of beliefs (a set of beliefs that reveal that she has not properly grasped the concepts she is using). This fact gives us reason to conclude that such propositions have moral content. However, for this objection

to succeed, it must be possible to coherently accept any of the four analyses discussed above and yet deny that such acceptance has any implications for action.

This objection initially seems appealing because it *does* seem like it would be coherent for someone to believe one of the theories discussed above yet also reject the idea that moral considerations have any implications for action. For example, it seems coherent to acknowledge that moral discourse is committed to a faulty conception of mind-independent, objective moral values, yet continue to use a non-moral procedure in order to figure out what to do. I might, for example, make judgments about what I ought practically to do based entirely on what I prefer without paying any attention to moral considerations.

In other words, this argument presumes that unless we agree to the following proposition:

IMP: Moral considerations have implications for action.

we cannot conclude that the four theories discussed above have any implications for action. We could, for example, adopt an analysis of the nature and content of moral judgment without agreeing that the outcome of moral deliberation is even remotely relevant to practical reasoning. This attitude towards morality would not be novel. Both Nietzsche (1887) and Williams (1985) proposed theories about the content and psychological nature of moral judgments. Both authors then denied that moral judgments

were relevant to figuring out what to do or how to live. Since it seems entirely consistent for us to accept that morality is best understood in a particular way (e.g., as Williams' 'the morality system' or Nietzsche's 'slave-morality') and to simultaneously reject that we should take morality to give us reasons for action, it is plausible that the theories of moral judgment I discussed above do not have implications for action. If IMP is denied, then we could accept a theory about the nature of moral judgment without being committed to any propositions that have implications for action. Our place would be very much like that of Nietzsche or Williams after respectively rejecting slave-morality and the morality system. Despite the fact that they recognize that certain things are relevant to moral thought, they deny that moral thought has any bearing on how, practically, we ought to live. Once we reject morality, we would then be free to pursue the project of determining how to live.

The objection under consideration is therefore successful up to this point, as it has shown that it is not a conceptual truth that the four theories of moral judgment discussed above have implications for action. However, this objection is also committed to the claim that propositions like A *do* have implications for action. Insofar as this objection attempts to differentiate moral propositions from non-moral propositions on the grounds that the former have implications for action and the latter do not, proponents of this view must either concede that A has implications for action or that A is not a moral proposition. If this objection is committed to the claim that A lacks moral content, then the objection must be rejected because A is a paradigmatic example of a moral proposition. In order to show that A has implications for action, this objection must be committed to the claim that IMP is somehow built into A or implied by A in such a way that it is not coherent to

accept A but deny IMP. So, to summarize, this objection is committed to the following two claims:

1) It is coherent to accept the four theories discussed above while denying IMP.
2) It is not coherent to accept A while denying IMP.

This is the only way that we can understand the claim that A has implications for action in a way that the four theories discussed above do not. The fact that there is this distinction between the class of paradigmatically moral propositions and theories about the nature of moral judgment suggests that there is a wide gulf between the two which I have failed to recognize.

My response to this argument rests on the claim that propositions such as A and normative moral standards such as the following:

> UTIL: An act is morally permissible if and only if it maximizes the amount of goodness in the universe.

have implications for action in the *same circumstances* as do the theories of moral judgment discussed above. This is because it is coherent for someone to accept UTIL or A but also claim that considerations about moral permissibility do not bear on considerations about how to live and thus do not have any implications for action. In other words, I deny that (2) is true insofar as one can coherently deny IMP for propositions like A and UTIL as well as for theories about the nature of moral judgment.

There are many ways that we can go about highlighting the fact that one can accept that some paradigmatically moral propositions are true yet deny that such propositions have any implications for action. For example, it is a plausible assumption that the proper criterion for adjudicating between normative moral theories is their fit with our moral intuitions. Alternatively, we might endorse a conception of what Norman Daniels (1979) has called wide reflective equilibrium and thus also bring certain background considerations (about, for example, the nature of a person) to bear on our theory choice. I might use either of these methods to argue in favour of UTIL as the best of a set of possible normative moral theories but also disagree with the claim that normative moral theories are at all relevant to determining what to do. While this might be a strange position for one to adopt, it would not be an incoherent one.

For the remainder of this chapter, I will argue that there are no plausible arguments in favour of the claim it would be incoherent to believe that paradigmatically moral propositions like A and UTIL do not have any implications for action. I also argue that despite this fact, there are substantive reasons to believe IMP and thus to think that moral considerations do give us reasons to act certain ways. Furthermore, the same substantive considerations that demonstrate that the class of paradigmatic moral propositions have implications for action also show that IMP is true for theories about the nature of moral judgment. This conclusion therefore undermines the argument under discussion, as it shows that propositions like A and UTIL have implications for action to the same extent as do the four theories discussed above. I believe that this conclusion is most clearly supported by what some philosophers have said about the metaethical position known as internalism. In order to show that there are substantive reasons to think

that the four theories discussed above have implications for action, I will now turn to a discussion of several ways of understanding internalism.

'Internalism' is a term used to denote a plethora of positions, all of which assert that there is some connection between reasons, moral judgments and motivating attitudes. There are a few versions of internalism that might warrant the claim that all denials of IMP are incoherent. One of these versions of internalism amounts to the claim that "moral obligations are, or entail, practical reasons" (Shafer-Landau 2003, p. 192). This view is sometimes called 'moral rationalism' (Shafer-Landau 2003) and sometimes called 'internalism about reasons' (Brink 1989 p. 39). Sergio Tenenbaum has articulated the clearest statement of this form of internalism. Tenenbaum attributes one of the following two assumptions to this kind of internalism:

> *Reason-Giving (Judgment) Assumption*: If one accepts that X is morally right, one thereby accepts that there is a reason to do X.
>
> *Reason-Giving (Existence) Assumption:* If X is morally right, then there is a reason to do X. (Tenenbaum 2000 p. 109)

These theses are differentiated by the fact that the first (*Judgment*) makes a claim about what one is committed to upon making a moral judgment, and the latter (*Existence*) makes a claim about what follows from the truth of a moral fact. Tenenbaum considers the disjunction of these two to best represent this kind of internalism. For the sake of clarity and brevity, I will refer to the view that treats the disjunction of the reason-giving assumptions as true as INTERNALISM.

Some arguments try to establish INTERNALISM as a conceptual truth by attempting to demonstrate that if one were to deny INTERNALISM, one would be saying something incoherent. If this were the case, then it would turn out that IMP is also a conceptual truth, as INTERNALISM implies it. However, both IMP and INTERNALISM are better understood as substantive truths than as conceptual truths. Furthermore, the substantive reasons for endorsing INTERNALISM commit us to accepting IMP for *all* moral propositions, including theories about the nature of moral judgment.

I will now consider some plausible arguments that INTERNALISM is a conceptual truth, and conclude that they are unsatisfactory. I will then offer some arguments in favour of the claim that INTERNALISM is a substantive truth (or, in other words, a proposition that is true but that it would be coherent to deny), rather than conceptual truth.

One plausible argument that INTERNALISM is a conceptual truth might propose that ethical thought just *is* deliberation about what to do. Christine Korsgaard, for example, implies such a view when she writes the following: “if you think reasons and values are unreal, go and make a choice, and you will change your mind” (Korsgaard 1996 p. 125). Note, however, that the term ‘morality’ as it is used in IMP is discussing narrowly moral considerations, not broadly moral considerations. When we identify morality with practical reasoning, as does Korsgaard, we shift to a discussion of moral practice understood broadly. It is true that, broadly, moral reasons are just reasons for actions, but narrowly, it still seems coherent to be in the position of Nietzsche or Williams and accept an understanding of morality (construed narrowly) while denying that the system of morality is at all relevant to determining what to do (or, in other words, has broadly moral

implications). The type of argument that follows from Korsgaard's claim would, if successful, imply that INTERNALISM is a conceptual truth. However, this argument does not succeed because it conflates the narrow and broad uses of the term 'morality.'

Another argument that INTERNALISM is a conceptual truth attempts to derive INTERNALISM from other conceptual truths. For example, such an argument might begin by claiming that the following two propositions are conceptual truths:

1) Moral properties motivate[24]
2) Motivation is sufficient to provide a reason for action [25]

The argument could then conclude with the claim that INTERNALISM follows from these two premises. To be clear, this argument is relying on the claim that anyone who denies that moral properties (e.g. moral wrongness) provide individuals any motivation to do anything is saying something incoherent. Similarly, this argument is claiming that anyone who denies that being motivated to do X gives us reason to do X is also saying something incoherent.

If successful, this argument would establish that INTERNALISM is a conceptual truth. However, these premises are very controversial and there are some plausible arguments against them in contemporary philosophical literature. Consider, for example, David Brink's arguments against motivational internalism. Motivational internalism is the view that either the existence of moral properties or the recognition of moral

[24] This view is one form of moral internalism, often referred to as motivational internalism and sometimes thought to be a conceptual truth (See Brink 1989 for an elaboration of the details of this view).
[25] Williams's (1981) view of the nature of reasons implies this.

properties motivates people to act in certain ways. This view is basically equivalent to (1) and should not conflated with INTERNALISM.

Brink's argument is that motivational internalism does not take amoral skepticism (moral skepticism that questions whether moral considerations are sufficient to motivate people) seriously insofar as it treats it as nothing more than a conceptual confusion (Brink 1989). Brink is correct, I think, in claiming that the amoralist skeptic is offering a genuine, substantive challenge to moral practice, and any theory that rejects amoralist skepticism as incoherent is not taking this challenge sufficiently seriously. Regardless of whether or not Brink's argument is in fact successful, his argument is sufficiently plausible to raise doubt that (1) is a conceptual truth.

There are also reasons to reject (2). If there are any instances wherein we are motivated to do X but also think that we do not have a reason to do X, such an instance would serve as a counter-example to (2). It is easy to imagine someone feeling an urge to do something that would not fulfill any goal they currently have and would substantially frustrate their interests as well as the interests of those close them. Such an individual is clearly motivated to do the action at hand, but it seems plausible that, nonetheless, this person has no reason to do it.[26] The plausibility of cases like this suggests that (2) is false and if (2) is false, it certainly cannot be a conceptual truth.

Even if we do not wish to say that (2) is false, there is still reason to deny that it is a conceptual truth. Many philosophers have distinguished motivating from normative reasons (Parfit 2011 p. 66; Scanlon 1998 p. 19) and in doing so have claimed that we can

[26] See Parfit 2011 chpt. 3 for an argument to this effect.

be motivated to do X (we can have a motivating reason to do X) without having any normative reason to do so. Even if these arguments do not succeed, it does not seem like these philosophers have said anything incoherent in making these assertions. But if (2) is a conceptual truth, then these philosophical positions must be incoherent. If (2) is true, it must be because substantive reasons count in favour of it, rather than because it is a conceptual truth.

Neither of the arguments I have here addressed successfully demonstrate that INTERNALISM is a conceptual truth. Further evidence that it is not conceptually true follows from the fact that we seem to be able to meaningfully argue with people who deny it. Consider one such person, a 'reasons amoralist' who recognizes the truth of moral facts, but denies that these truths give any reason for action (See Brink 1989 pp. 46-47 for an elaboration of this view). This individual seems to be asking a perfectly reasonable question of the proponent of INTERNALISM: why should I understand morality as giving me reasons to behave one way or another? If INTERNALISM is a conceptual truth, this question would have to be empty or incoherent. But there is a long philosophical tradition of asking this question. As Tenenbaum points out, it is practically identical to that posed to Plato by Glaucon (Tenenbaum 2002 p. 112).

The challenge to the individual who holds that INTERNALISM is a conceptual truth is to show that all the philosophers who have assumed that it is coherent to deny INTERNALISM are mistaken. I have yet to see an argument that satisfactorily accomplishes this goal. There is therefore insufficient evidence to show that INTERNALISM is a conceptual truth, and we must therefore proceed with the assumption that it is not. And if INTERNALISM is not a conceptual truth, then the argument I outlined at the beginning of

this section (which, you will recall, claims that there is a rigid distinction between paradigmatically moral propositions and theories about moral judgment regarding their implications for action), does not succeed.

Despite the fact that INTERNALISM is not a conceptual truth, I believe that there are *substantive*, not conceptual, arguments that establish its truth. Furthermore, these arguments imply this conclusion for *both* paradigmatically moral propositions and theories about the nature of moral judgment. This further undermines the objection under discussion. For the remainder of this section, I will discuss and defend some of these arguments.

One argument in favour of INTERNALISM as a substantive truth is put forward by Shafer-Landau. This argument turns on the claim that those individuals who deny INTERNALISM (or what Shafer-Landau calls moral rationalism) are committed to a morally repugnant moral standard that is unfair. Regarding INTERNALISM, Shafer-Landau writes that:

> It seems unfair to criticize violations of such standards while admitting that an agent responsible for offensive conduct may have had no reason to do otherwise. The fairness and appropriateness of moral evaluation rest on an agent's attentiveness to reasons. An agent who correctly claims to have ignored no reasons for action cannot be held to have violated any moral standard. This plausible thought is only true if moral rationalism is true. (Shafer-Landau 2003, p. 193)

This argument amounts to the claim that INTERNALISM (or, as he calls it, moral rationalism) must be true for substantive moral reasons, rather than conceptual reasons.[27] The substantive moral reason in question is that it would be unfair to say that someone who had no reason to abstain from doing something should be blamed for doing it.

This is an interesting argument, but I think it succumbs to an important objection, namely, that it is circular. If the reasons amoralist is questioning whether or not moral considerations provide reasons, then presumably she would also question the claim that the moral reason provided above gives her any reason to believe INTERNALISM. If she does not care about morality, why should she care about fairness? I think that this is a legitimate concern, but we can salvage this argument by reframing Shafer-Landau's response as a claim about what is at the core of our moral thinking. By denying that morality gives us reasons for action, we come to be committed to a very counterintuitive picture of morality according to which we blame people for doing that which they have most reason to do. Morality, under this picture, becomes so unfamiliar that it is hard to understand how this could be a correct analysis of the relationship between moral considerations and reasons for action. The fact that moral systems that deny INTERNALISM are so counterintuitive gives us substantive reason to think that INTERNALISM is true.

Tenenbaum also offers an argument that gives us substantive reason to think that INTERNALISM is true. Rather than giving us moral reasons for endorsing INTERNALISM, he gives us pragmatic reasons for doing so. As he puts it, "morality would certainly lose its point if we no longer accepted that it gave us reasons for actions" (Tenenbaum 2002 p. 121). Unless we grant INTERNALISM, in other words, our moral practices would become

[27] Although Shafer-Landau also endorses the conceptual argument (Shafer-Landau 2003 p. 192).

useless, and we would no longer have any reason to engage in moral deliberation. This argument does not claim reasons amoralists are committed to incoherent or morally repugnant theses, but rather claims that such amoralists have simply failed to take notice of why morality is useful for us. As the reasons amoralist understands it, morality is an austere practice that has nothing to do with our reasons for action. But the amoralist has not recognized the fact that morality serves an essential purpose: it allows us to discuss, in a simple direct manner, how we should think and feel about certain types of people and actions. It makes it possible for us to coordinate our responses to many features of human behaviour, and it provides us a framework to praise people and blame them for their actions. If we do not understand moral deliberation as deliberation about what reasons we have to act and feel in certain ways, it could not perform this valuable function. Any understanding of morality that does not understand moral deliberation as giving us reasons therefore misses the point of engaging in moral deliberation. The pragmatic utility of understanding moral considerations as giving us reasons for action therefore gives us a substantive reason to endorse INTERNALISM. Furthermore, there is no non-arbitrary reason to limit this conclusion to only the class of paradigmatically moral propositions. If this argument succeeds for propositions like A and UTIL*,* it also demonstrates that the four theories discussed above have implications for action. After all, all four of these theories give us moral advice in at least some possible situations, and therefore indicate what reasons we have for acting in those situations.

Some philosophers would respond to these substantive arguments by claiming that moral considerations only give us reasons for action insofar as morality contingently helps us accomplish things that matter to us (Railton 1986; Brink 1989). This might serve

as a challenge to the argument I have just rehearsed insofar as the objection construes moral propositions as only contingently reason-giving. However, such an objection would not serve to refute my argument that the reason-giving assumption is true, as the assumption in question says nothing about *why* moral judgments give us reasons for action. What is important is that deliberation about reasons is a central component of moral thought insofar as the falsity of INTERNALISM would commit us to an unrecognizable and pointless conception of morality. The fact that moral considerations give us reasons only contingently does not serve to refute my claim that these considerations do in fact give us reasons.

To summarize, I began this section with the concern that it intuitively seems like paradigmatically moral propositions like A and UTIL give us reasons for action, but it does not seem like theories about the nature of moral judgment give us such reasons. However, since I claim that all of these propositions have moral content, I must respond to this concern. In service of this goal, I have argued that it is not a conceptual truth that moral considerations have any implications for action. However, some prominent work on the metaethical position I have dubbed INTERNALISM suggests that there are *substantive* reasons to think that moral considerations give us reasons to act in specific ways. Furthermore, these arguments extend to *all* propositions that have moral content, including the four theories discussed above.

IV - Conclusion

In this chapter, I have argued that some prominent metaethicists are approaching a particular question with some mistaken assumptions. Specifically, their attempts to articulate and defend a theory about the nature of moral judgment conflate propositions that have moral content with propositions that lack moral content. This conclusion has profound implications for metaethical methodology. Some of these consequences are the central topic for the third and final chapter of this dissertation. In the next chapter, I discuss the consequences of this argument for recent attempts to study the nature of morality by appealing to recent findings in neuroscience, experimental psychology and evolutionary theory.

Chapter Three

Theories of Moral Judgment Are Not Empirical Hypotheses

My thesis in the previous chapter was that a number of philosophers have misunderstood the distinction between normative ethics and metaethics. Most philosophers interpret claims about the nature and content of moral judgments as morally neutral, psychological theories. This is an intuitive and common view. After all, it seems like these theories are describing contingent features of human psychology. However, in many cases, theories about the nature of moral judgments must be understood as having moral content and thus as being internal to moral discourse. This implies, among other things, that these theories must conform to our considered moral judgments and commitments and can be refuted by demonstrating that they are inconsistent with true moral propositions. Whatever methodology we decide upon in order to confirm or disconfirm moral propositions must also be used to confirm or disconfirm theories about the nature of moral judgment. Additionally, if moral propositions cannot be confirmed or disconfirmed (i.e., if it turns out that moral knowledge is unattainable), the same must be true for these theories of moral judgment.

In this chapter, I will argue that any theory about the psychological nature of moral judgments that has moral content cannot be understood as an empirical hypothesis. Many philosophers have recently tried to defend theories about the nature of moral judgment by appealing to evidence from the empirical sciences of psychology and neuroscience. I argue that these philosophers conflate moral propositions with non-moral

ones and predicate their putatively empirically-informed arguments on substantive moral assumptions. The consequence of this is that these theories cannot be articulated in a morally neutral manner and thus cannot be understood as empirically-confirmable hypotheses. As an alternative to this methodology, I propose that the adjudication of theories about the nature of moral judgment must be predicated on a recognition of the fact that these theories are fundamentally grounded in the moral domain. This implies, among other things, that moral considerations provide the primary body of evidence for or against these theories.

I - Empirical Hypotheses

First, I should explain how I understand the distinction between moral propositions and empirically-confirmable scientific hypotheses. My discussion of empirical hypotheses is motivated *only* by the fact that many metaethicists have attempted to confirm or disconfirm theories about the nature and content of moral judgment by appealing to evidence from empirical sciences. I am therefore interested only in how these philosophers understand the nature of empirical evidence. An in-depth account of the nature of empirical evidence or empirical hypotheses that would address all relevant questions from epistemology and the philosophy of science would therefore take me far beyond the scope of this chapter. For my purposes, I will use the term ‘empirically-confirmable hypothesis’ and related terms in a manner that is consistent with how these philosophers use them. Specifically, I will focus on three philosophers who are prominent proponents of empirically-informed metaethics: Shaun Nichols, Jesse Prinz

and Richard Joyce. Nichols endorses a form of empirically-informed metaethics that pays attention to "controlled experiments in moral psychology" performed by social scientists (specifically psychologists) in order to identify the "empirical details" that will help philosophers give adequate answers to questions about "the nature of moral judgment ranged widely, including issues about the role of sentiment in moral judgment, the role of reason in moral judgment, and the origin of moral judgment" (Nichols 2004 p. 4). These comments may not amount to a rigorous definition of the term 'empirically-confirmable hypothesis' or any related terms, but they do indicate what Nichols has in mind when he advocates empirically-minded philosophy. Empirical hypotheses, according to Nichols, are those that are confirmed or disconfirmed by the performance of controlled experiments of the sort performed by psychologists and other natural and social scientists. Many metaethical debates can, on Nichols' view, be settled by investigating how individuals respond to certain questions about the nature and status of their moral judgments. He cites Turiel, Nucci, Kohlberg and Piaget as paradigmatic examples of psychologists who used this methodology (ibid pp. 4-5).

Prinz also argues in favour of the use of empirical methods to inform metaethical debates. He develops and defends his own detailed, empirically-informed account of the nature of moral judgment (Prinz 2007a, 2007b, 2008) which will be discussed below. Like Nichols, Prinz believes that "empirically minded philosophy" is best understood as philosophy in which "philosophers cite laboratory studies in support of their theories" (Prinz 2007a p. 271). This suggests that Prinz, like other philosophers in this area of research, understands empirically-confirmable hypotheses as theses that can be confirmed

or disconfirmed (or at the very least, informed) by the kind of experimental work that is typical of psychology.

Like Nichols and Prinz, Joyce does not provide a rigorous definition of the term ‘empirically-confirmable hypothesis’ or any related term, but he does offer an account of what kinds of experiments could serve to answer questions about the phenomenology of moral judgment. Joyce endorses the methodological approach of determining answers to questions about the phenomenology of moral judgments by giving subjects surveys on the topic. He goes on to say that “data must be collected in an intellectually responsible manner, complying with the customary scientific standards concerning such things as sample size, control groups, replicability, randomization, correcting for framing effects, and so on” (Joyce 2009, p. 68). This indicates that when Joyce advocates understanding a theory as an empirical hypothesis, he too believes that this implies treating that theory as if it can be confirmed or disconfirmed by engaging in the kind of rigorous experimental process that is typical of psychological studies and research performed by other social and natural scientists.

These considerations do not provide us with a rigorous answer to the question I posed above about the nature of empirical hypotheses. However, this survey does give us enough information for my stated purpose. In order to be responsive to Joyce, Prinz and Nichols I will assume an account of empirical hypotheses that is *disciplinarian:* a proposition is best construed as an empirical hypothesis if and only if it can be confirmed or disconfirmed by the kinds of experiments performed by social scientists and natural scientists. This definition best captures how Nichols, Prinz and Joyce use the term

'empirically-confirmable hypothesis' and it conforms with the usage of many of the metaethicists that I will discuss below.[28]

In the remainder of this chapter, I will argue that theories about the nature of moral judgment cannot be understood as empirically-confirmable hypotheses. This is because these theories have moral content. The following two sections of the chapter canvass and assess a number of arguments in favour of the claim that any proposition that has moral content is not identical to or reducible to any empirically-confirmable hypothesis. I then apply those considerations to a discussion of Nichols, Prinz, and Joyce.

II - Non-Naturalism and Ethics as a Branch of Philosophy

Many philosophers endorse the view that substantive moral propositions are in some way distinct from the class of propositions that can be confirmed or disconfirmed by empirical investigations. Sometimes the arguments in favour of this position take the form of the claim that ethics cannot be understood as a naturalistic discipline. This thesis can be cashed out in many ways. One way of doing so is to align oneself with philosophers who are committed to a form of moral non-naturalism that understands moral properties as both instantiated and not natural.

This form of non-naturalism is, however, fairly vague and open to a number of criticisms. One major issue is that it is not entirely clear where we should place the borders of the naturalistic domain or, in other words, what it means for a property to be

[28] See my discussion of Shafer-Landau, Moore, and Gibbard in the following sections.

natural. Russ Shafer-Landau, a prominent proponent of moral non-naturalism, has recognized the complexity of this question and has advocated a simplified approach. He believes that the moral non-naturalist's thesis is most clearly understood if we conceive of the natural as "whatever is the object of study by the natural sciences" (Shafer-Landau 2006 p. 211). This disciplinary conception would therefore identify something as natural "just in case it figures ineliminably in true propositions that emerge from (on some accounts, the perfected version of) physics, chemistry, molecular biology, astronomy, etc. Leave astrology off that list. Film studies too" (Shafer-Landau 2006 p. 211). This approach clarifies what exactly the non-naturalist is trying to say about moral properties, namely, that they are not the sorts of thing that are or could be the object of study of any known empirical discipline. This quotation from Shafer-Landau demonstrates that non-naturalists understand 'naturalistic propositions' in roughly the same way that I defined 'empirically-confirmable hypotheses' above.

Shafer-Landau's own judgment regarding the distinction is as follows:

> I don't think that ethics is a natural science. Its fundamental principles are not inductive generalizations. It is not primarily concerned with causal efficacy. Its central principles are not descriptive of historical contingencies. The phenomena it does describe are supervenient as a matter of conceptual requirement. It allows for a much greater degree of indeterminacy and vagueness than is found in typical natural sciences. It has only a very little concern for mathematical quantification and precision. Unlike any of the recognized sciences, its truths are normative truths that direct and guide, rather than (in the first instance) predict the course of future events or explain what has already occurred. Moral truths provide justifying reasons that are often ignored. Physics and geology and hydrology don't do that. (Shafer-Landau 2006 p. 211)

This powerful conclusion on Shafer-Landau's part is dependent upon his understanding of ethics as a part of philosophy. As he does for natural sciences, he describes philosophy in a disciplinary manner: philosophy is whatever philosophers do. The most important feature of philosophical practice that Shafer-Landau emphasizes is that, in contrast to the *a posteriori* natural sciences, philosophy is a paradigmatic example of an *a priori* discipline. According to Shafer-Landau, the status of philosophy as an *a priori* discipline comparable to mathematics disqualifies it from being considered a natural science. This precludes the possibility of assessing philosophical claims from the perspective of one of the natural scientific disciplines canvassed above. The fact that "ethics is a branch of philosophy"[29] leads Shafer-Landau to the conclusion that "the conditions under which actions are right, and motives and characters good, aren't confirmed by the folks with lab coats. They are confirmed, if at all, by those who think philosophically" (ibid, p. 217).

While I find Shafer-Landau 's arguments persuasive, I do not wish to rely too heavily on these considerations alone. It is not obvious how we ought to understand the nature of philosophy and the relationship between philosophical work and the natural sciences. Many of the philosophers I will be discussing below would likely reject some aspects of Shafer-Landau's assumptions about the *a priori* status of philosophy, so beginning with the assumption that Shafer-Landau is correct may lead me to beg the question against the theories I will discuss.

My goal is to offer an argument to this effect that will allow me to apply similar considerations to Nichols, Prinz and Joyce's work without the threat of circularity. In the

[29] Shafer-Landau uses the term 'ethics' in roughly the way I have been using the term 'moral philosophy.' I treat these terms as interchangeable.

following section, I will survey some other arguments in favour of the claim that moral propositions cannot be construed as empirically-confirmable hypotheses and will identify what I consider to be the strongest of these arguments (the so-called "What's at issue?" argument). This argument will draw on a set of metaethical considerations that are indebted to G.E. Moore's arguments regarding the so-called "naturalistic fallacy." I will then turn to the work of Nichols, Prinz and Joyce in order to demonstrate that their theories are either predicated upon substantive moral presuppositions that they have not recognized or that they mistakenly conflate moral propositions with empirically-confirmable hypotheses. A consequence of these arguments is that the theories proposed by Joyce, Nichols and Prinz about the nature of moral judgment are at bottom dependent upon normative moral propositions, and therefore are grounded in moral discourse rather than empirical science.

III - Non-Natural Concepts and Moorean Considerations

The belief that moral philosophy must be distinguished from science can be severed from the kind of non-naturalism found in Shafer-Landau's work. Rather than endorsing the claim that there are instantiated non-natural moral properties, some philosophers hold that the only properties that are instantiated are natural but that morality nonetheless resists reduction to natural sciences because of the distinctively normative nature of our moral concepts. One example of such a view is found in Allan Gibbard's work:

> Whereas the concept of being good is distinct from any naturalistic concept – from concepts fit for empirical science and its everyday counterparts – the property of being good is a natural property, a property for which we could have a naturalistic concept. (Gibbard 2006a p. 323)

Gibbard's positive theory is that normative concepts behave very much like naturalistic concepts on the surface, but are distinct from naturalistic concepts insofar as they serve to express mental states that are best understood as being like plans (Gibbard 2003). The details of Gibbard's plan-expressivism are not relevant for my purposes. I am interested solely in the arguments he and others put forward to establish that moral concepts resist reduction to naturalistic concepts. If this can be demonstrated, it would imply that moral propositions cannot be understood as empirical hypotheses. The arguments I will discuss below are indebted to G.E. Moore, but the strongest versions of these arguments substantially diverge from Moore's own. Hence, I refer to this body of arguments as 'Moorean considerations' rather than simply as Moore's arguments.

For the purposes of my discussion of the Moorean considerations I will be adopting two conventions. First, I will continue to follow philosophers such as Moore, Gibbard, and Shafer-Landau in construing the debate about the role of empirical sciences in metaethics as being closely related to debates about whether moral properties and concepts should be construed as naturalistic or non-naturalistic. I do not, however, wish to enter into debates about the various forms of naturalism in metaethics and philosophy more broadly.[30] In the arguments below, I will use the term 'natural' in the way that I described above: whenever a property or concept is natural, it does or could fall within the purview of one of the empirical, natural or social scientific disciplines that Shafer-

[30] See Miller 2003 chapters 8 and 9 for a detailed account of different forms of metaethical naturalism.

Landau identified above. When I consider the possibility of a moral term having a 'naturalistic analysis,' I am, in effect, considering whether or not propositions containing that term can be understood as empirically-confirmable hypotheses.

The second convention I will adopt is to use the word 'good' as an example in many of the arguments that follow. This is because I wish for my arguments to be continuous with Moore's and for his influence on them to be explicit. Since Moore believed that the term 'good' was a simple, indefinable notion, the term played a central role in his arguments for non-naturalism. While I often use 'good' in the same contexts as Moore, I wish to be explicit that the term, as I use it, is meant to stand for a generic moral predicate and that any of my arguments that contain it could be coherently re-articulated with a different moral predicate. I am not committed to any of Moore's claims about the indefinability or simplicity of the term 'good.'

I now turn to the arguments themselves. Moore's arguments that moral terms do not submit to naturalistic analyses are situated in the context of his discussion of the naturalistic fallacy. While not actually a logical fallacy or a mistake in deductive reasoning, the error Moore identifies is alleged to be present in many philosophical treatises. Most succinctly, the naturalistic fallacy is committed whenever someone conflates substantive moral propositions about which things are good with a definition of the term 'good' (Moore 1903 p. 62). We might, for example, concede that kindness, knowledge and pleasure are all good. We may even go as far as to claim that these things are exhaustive of the category of good things. However, we commit Moore's naturalistic fallacy whenever we conflate these types of claims about the substance of morality with definitions of moral terms.

William Frankena (1939) argues that the naturalistic fallacy is actually not unique to morality, but is rather an instance of a more widespread mistake, the definist fallacy:

> To judge by the passages I have just quoted, the definist fallacy is the process of confusing or identifying two properties, of defining one property by another, or of substituting one property for another. Furthermore, the fallacy is always simply that two properties are being treated as one, and it is irrelevant, if it be the case, that one of them is natural or non-ethical and the other non-natural or ethical. (Frankena 1939 p. 471)

According to Frankena, the naturalistic fallacy in ethics is just a specific example of a mistake that occurs in many areas of philosophy. More importantly, Frankena points out that Moore's accusations regarding the naturalistic fallacy do not constitute an argument in favour of the claim that moral terms do not submit to naturalistic analyses, but must rather be treated as the conclusion of an argument to that effect. Moore's attempt to provide such an argument relies on the claim that moral definitions (specifically, the definition of the word 'good') remain 'open' in a way that the definitions of non-moral terms do not. A definition remaining 'open' in Moore's sense means that it is always coherent and significant[31] to question the definition without error or confusion. The argument Moore put forward is usually referred to as the 'Open Question Argument.'

Moore's example of the phenomenon described by the Open Question Argument is the proposed naturalistic analysis:

[31] 'Significant' is Moore's term. I believe it can be faithfully rendered to be consistent with how I have been and will continue to use the terms 'coherently' and 'meaningfully.' I will use these terms as being roughly interchangeable for the remainder of this section.

> N: When we think that A is good, we are thinking that A is one of the things which we desire to desire. (Moore 1903 p. 67)

The open question argument turns on the presupposition that whenever readers entertain N they will find that they can still coherently ask themselves whether or not it is always true that those things which we desire to desire are good. N therefore cannot be an accurate analysis of the term ‘good’ because the question of what is good always seems to be ‘open.’ No similar analysis of the term ‘good’ is sufficient because we can always significantly ask whether or not that definition will serve as an accurate moral standard. If the definition were correct, then the question would become ‘closed’ and we could no longer coherently ask it. For example, if N were a correct definition, then competent English speakers could not coherently ask whether or not all things which we desire to desire are good. However, since competent English speakers can coherently and significantly ask this question, N cannot be a correct analysis (although it may be substantively true that all and only that which we desire to desire is good). Further, Moore believes that the same would be true for any definition of the term ‘good.’

However, Moore argues that the openness phenomenon does not occur for naturalistic terms. He demonstrates this by considering the term ‘horse.’ We can, according to Moore, define the term ‘horse’ by enumerating all the properties and qualities of being a horse (Moore 1903 p. 59). Once we have identified an individual creature as having all of those properties, we can no longer significantly question whether or not it is a horse. As we saw above, Moore claims that this is not true of moral terms (specifically the term ‘good’). Proposed moral definitions are therefore open in a sense that definitions of naturalistic terms are not.

This argument is problematic because Moore's understanding of the reductive definitions of complex ideas (according to which the definitions of such terms is simply a reduction to a number of simple ideas) is untenable. For example, contemporary philosophers recognize that it is quite difficult to formulate a rigorous, uncontroversial definition of what it means for an individual to be a member of a species.[32] Moore's original statement of the Open Question Argument is therefore untenable, as the 'openness' of moral questions can be compared to the openness of many naturalistic questions as well. Richard Joyce, for example, writes that the Open Question Argument is "based on the confused views of necessity, a prioricity, and analyticity that dogged early-twentieth-century philosophy and weren't straightened out until the middle of the century" (Joyce 2006 p. 152). Because of Moore's confused views on these matters, he fails to notice that the openness he identifies is not unique to moral questions. Joyce's preferred example revolves around an argument regarding whether or not water is H2O. An individual with a naïve conception of chemical theory may coherently and meaningfully question whether or not water is H2O. This individual's mistake is that he has endorsed an inadequate chemical theory. His question is still coherent and meaningful, even though he is wrong for substantive reasons. There is a sense, then, in which many (if not all) complex naturalistic definitions remain open in Moore's sense.

Nonetheless, the idea that definitions of moral terms are open in a sense that naturalistic definitions are not remains persuasive. Some contemporary philosophers have attempted to explain this 'open feel' by emphasizing the distinctively practical character of moral judgments. Darwall, Gibbard and Railton (1992) have attempted to explain the

[32] For a succinct summary of the debate regarding species in the philosophical literature, see Pigliucci and Kaplan 2006 pp. 214-217.

fact that one can always intelligibly question whether any analysis accurately captures what is good by arguing that moral claims have a "conceptual link with the guidance of action, a link exploited whenever we gloss the open question 'Is P really good?' as 'Is it clear that, other things equal, we really ought to, or must, devote ourselves to bringing about P?'" (Darwall et al. p. 117). This view explains the 'open feel' of moral questions by claiming that moral propositions (e.g., 'only pleasure is good') are best understood as being about what one ought to do (promote pleasure) and strive to bring about in particular circumstances (pleasure). This analysis relies on a thesis I discussed in chapter two, the reason-giving assumption:

> *Reason-Giving (Judgment) Assumption*: If one accepts that X is morally right, one thereby accepts that there is a reason to do X.
>
> *Reason-Giving (Existence) Assumption:* If X is morally right, then there is a reason to do X. (Tenenbaum, p. 109)

As I argued in chapter two, the reason-giving assumption (by which I, following Tenenbaum, mean the disjunction of these two assumptions) is true for substantive reasons but not, as some people have thought, for conceptual reasons. A person can, without confusion, deny the reason-giving assumption. However, if we do not accept this assumption, morality would lose much of its point and purpose. The point of moral discourse is to allow us to discuss and deliberate about the reasons that we have for performing certain actions and having certain attitudes. There would be very little reason to persist in a practice that had no connection to these reasons and there are therefore

good reasons to grant the reason-giving assumption. Furthermore, as Shafer-Landau has pointed out, a moral system that rejects the reason-giving assumption would be unrecognizable (Shafer-Landau 2003).

According to Darwall et al.'s argument, the fact that our moral practice is shaped by the reason-giving assumption either contributes to or is solely responsible for the open feel of moral propositions. If this argument is correct, then no moral proposition is identical to any empirically-confirmable hypothesis because one can accept any empirical hypothesis without accepting any reasons as counting in favour or against specific acts. Since the acceptance of moral propositions *does* commit us to the acceptance of there being such reasons, the two classes of propositions must be distinct.

While there is reason to think that this argument succeeds, its rhetorical force is limited. This is because some philosophers who defend naturalistic analyses of moral terms also deny the reason-giving assumption. One example of a reductive naturalist of whom this is true is Peter Railton.[33] Railton addresses the following version of the reason-giving assumption that he attributes to Hume: "the thesis that morality is practical, by which he (Hume) meant that if moral facts existed, they would necessarily provide a reason for moral action to all rational beings" (Railton 1986 pp. 166-167). Railton rejects this statement of the reason-giving assumption and attempts to explain the action-guiding nature of moral judgment by pointing out that adherence to moral standards helps advance certain of our ends (Railton 1986 p. 170). Since at least some advocates of the reduction of the moral to the natural reject the reason-giving assumption, it would be preferable if we could formulate our arguments against analyses of the moral in natural

---

[33] See also Brink 1989.

terms without relying on the reason-giving assumption as a premise. Fortunately, there are other, even stronger ways of interpreting and revising the Moorean considerations. The remainder of this section is concerned with articulating and defending one such argument.

Gibbard (2003) argues that the most persuasive of the arguments found in Moore is not the open question argument, but rather an argument that can be glossed from one of Moore's examples. He dubs this the 'What's at issue?' argument because the argument leads us to ask exactly that about a moral disagreement (Gibbard 2003 p. 23). This argument amounts to a demonstration of the fact that naturalstic analyses (e.g., a possible naturalistic analysis of 'good' might be "that good means the object of desire," (Moore 1903 p. 63)) of moral terms misconstrue the nature of moral disagreement. In other words, such analyses cannot accurately capture what is at issue between individuals engaged in moral disagreements. Gibbard's articulation of this argument focuses on a particular example of a moral disagreement over the following two propositions:

A: Only pleasure is good.

and

B: Not only pleasure is good.

This example is meant to show that no naturalistic definition of the term 'good' is satisfactory because no such analysis can explain what is at issue between two individuals arguing over A and B. For example, it is not the case that the word 'good' can be defined in the following way: "good means the object of desire" (Moore 1903 p. 63). If this were the correct definition of 'good', then the question "what is at issue in the disagreement over A and B?" would be answered with the claim that the individuals holding to these propositions are disagreeing over whether or not pleasure is the only object of desire. However, it is coherent for the individual arguing in favour of B to accept B and also accept that only pleasure is desired. In short, this particular naturalistic analysis of the term 'good' cannot capture what is at issue between two individuals disagreeing over A and B. Their disagreement is about moral goodness and this cannot be reduced to a disagreement about what is desired. Gibbard thinks, and I agree, that an argument of a similar form would serve to refute all proposed naturalistic analyses of the term 'good.' This last claim (that arguments of this style can be used to respond to any naturalistic analysis) implies that no proposition containing a moral predicate is equivalent to any proposition without a moral predicate and thus that moral propositions cannot be equivalent to empirically-confirmable hypotheses.

The force of this argument can be made more perspicuous by looking at the difference between disagreements over ascriptions of naturalistic predicates and disagreements over the ascriptions of moral predicates. If I were to argue with someone about, to follow Moore's example, whether or not something was a horse, my interlocutor and I could take certain steps to ensure that we do not talk past each other. For example, we might begin this debate by making sure that we agree on a complete definition of

what we mean by the term ‘horse.’ We can then enumerate all the empirical observations that we take to be relevant to determining whether or not this creature is a horse. Finally, we can make explicit the norms of belief-formation that we endorse regarding horses. [34] If we find no disagreement between us on any of these points and yet our argument persists, one of us must be confused. Either we are mistaken in thinking that we endorse the same norms of belief-formation, we are mistaken in thinking that we share a definition of the term ‘horse’ or we are mistaken in thinking that we agree about all the relevant empirical evidence. The only explanation of why the disagreement persists is that we are talking past each other in one of these three respects and that this conversation is therefore somehow dysfunctional. ‘What’s at issue’ between us must be either a claim about the meaning of a term, an empirical observation or a disagreement over the norms of belief-formation in empirical sciences.

The same is not true of moral disagreement. Imagine that my interlocutor and I are not arguing about whether or not a specific creature is a horse, but rather about whether or not a particular action is morally right. We have both read Moore and been convinced that “‘right’ does and can mean nothing but ‘cause of a good result’” (Moore

[34] I include this last element, the norms of belief-formation, because some critics of this type of argument have suggested that a failure to include this element is a weakness of other versions of this argument. Railton (1986), for example, rejects a version of the argument I am here supporting. Railton’s statement of the argument is as follows: “two individuals who differ in ultimate values could, without manifesting any rational defect, hold fast to their conflicting values in the face of any amount of argumentation or evidence” (Railton 1986 , p. 166). Railton’s counterargument turns on the claim that belief-formation is a practice that also requires some basic agreement about the values and norms that govern scientific inquiry. Two individuals could therefore disagree over the norms of belief-formation without manifesting any rational defect. A consequence Railton draws from this is that there is nothing distinctive about morality in this respect. However, my argument in this section does not preclude the possibility that belief-formation is normative and practical in the ways Railton suggests. The point I wish to emphasize is that any two individuals can agree on all definitions, empirical evidence *and* the norms that govern scientific belief-formation and yet still meaningfully disagree about a moral matter of fact, whereas the same is not true for disagreements about naturalistic facts. I will argue for this last claim in the paragraphs that follow.

1903 p. 196) and we are in agreement that 'good' is a simple and indefinable notion (ibid pp. 58-59). We have also agreed on a standard of belief-formation in ethics (for example, we may have both agreed that one should affirm a belief in a moral proposition only after one has established a narrow reflective equilibrium between one's particular moral judgments and one's moral standards). Furthermore, we are in agreement about every relevant empirical observation about this situation. We have spent many hours detailing every non-moral fact that either of us takes to be morally relevant in this circumstance and we can find no disagreement there. Let us also stipulate that we are correct in this assessment. And yet we still disagree about whether or not the action is morally right. Unlike in the case of the disagreement about the controversial horse, this disagreement cannot *only* be explained by a misunderstanding between us or a disagreement about different norms of belief-formation (although either of these may well be the culprit in many moral disagreements). There is another plausible interpretation of this situation, namely, that we are engaged in a substantive moral debate about whether the action is wrong. 'What's at issue' between us in this case could either be a disagreement over meaning, empirical observations, the norms of belief formation in ethics *or* the status of a substantive moral proposition.[35]

There is therefore something at issue in moral disagreements that is not and cannot be at issue in disagreements about empirically-confirmable hypotheses, namely, the status of substantive moral propositions. These disagreements are distinctively moral

[35] The recognition of this feature of moral disagreements has a long history and is not unique to Moore and Gibbard. Stephen Toulmin (1950), for example, writes that "even if there is neither deception nor defect on either side, even if both parties are fully informed about the case and both mean the same by 'good' and 'right', it still makes sense to inquire whether their moral judgments are in fact the same." (Toulmin 1950 p. 20) I do not rely on Toulmin's argument, however, as he takes it to imply that values cannot be thought of as properties. I, however, think that we can meaningfully talk about moral properties in the sense described in Gibbard 2006a.

in a way that a disagreement over an empirical hypothesis cannot be. This fact provides the best evidence for the claim that it is impossible to reduce moral propositions to empirically-confirmable hypotheses. Disagreements about empirically-confirmable hypotheses must be explained as being about a definition, the norms of belief formation or an empirical fact. This exhausts the list of possibilities of what could be at issue in these cases. However, moral disagreements allow for an additional possibility, namely, a disagreement about the status of a substantive moral proposition. Any moral proposition can be meaningfully rejected or endorsed for moral reasons, whereas empirical hypotheses cannot be endorsed or rejected for these reasons. For example, the proposition 'murder is morally permissible' can be denied for the moral reason that murder is morally wrong. Furthermore, we can meaningfully disagree about whether we should accept or reject the proposition by engaging in a distinctively moral disagreement or, in other words, a disagreement over a moral proposition that cannot be explained as a disagreement over some non-moral matter. It is not, however, the case that we can meaningfully engage in a distinctively moral disagreement over whether or not we should accept the proposition 'Secretariat is a horse.' What is at issue here is either the definition of one of the words in the proposition, the norms of belief formation about what is or is not a horse, or some empirical observation about Secretariat. If someone were to claim that they agree with me regarding all three of these factors, yet persist in denying that Secretariat is a horse, this conversation would cause me to be baffled and confused (or, in other words, this conversation would not 'track' in this conversational context.[36] Moral reasons are simply not relevant to determining whether or not 'Secretariat is a horse' is true, whereas they are relevant to determining the status of propositions like 'killing in

[36] (I will explain my use of the term 'track' below.)

self defense is morally permissible.' In the case of this latter proposition, we could be arguing about definitions, norms of belief-formation or empirical observations, but we might also be having a brute disagreement about the moral status of killing in self-defense. There is, then, an additional dimension of disagreement that we can engage in with respect to moral propositions that we cannot engage in with respect to horse (and other naturalistic) propositions.

For the sake of clarity, here is a simplified version of the argument I have just rehearsed:

> P1: If we can meaningfully engage in distinctively moral disagreement (a disagreement over a moral proposition that cannot be explained as a disagreement over some non-moral matter, such as a definition), then moral propositions cannot be identical to or reduced to empirically-confirmable hypotheses
> P2: If distinctively moral disagreements track in conversational contexts, then we can meaningfully engage in distinctively moral disagreement
> P3: Distinctively moral disagreements track in conversational contexts
> C: Moral propositions cannot be identical to or reduced to empirically-confirmable hypotheses

P1 is true because if we can meaningfully engage in distinctively moral disagreement, this implies that we can reject or accept moral propositions for distinctively moral reasons (otherwise this type of disagreement would be confused and therefore would not be meaningful). However, since we cannot reject empirically-confirmable hypotheses for such reasons, moral propositions cannot be identical to or reducible to empirically-confirmable hypotheses.

P2 could potentially be rejected because of its reliance on a conception of what it means for something to track in a conversational context. It is true that I am relying to some extent on a fairly naïve and common-sense conception of what kinds of disagreements are coherent and meaningful, but I follow Gibbard in thinking that without assuming such a basic conception of what does or does not 'track' in a conversation or, alternatively, what types of statements should and should not cause bafflement in a conversational context, "no one could navigate a conversation. We couldn't even navigate our own inner thoughts" (Gibbard 2003 p. 26). In other words, I am relying on a common-sense conception of what does or does not 'track' in the context of a conversation, and I am presuming that the reader has an equal sense of when a statement does or does not make sense given relevant background information about the conversational to context (in most cases, at least.) As I am using these terms, whenever a disagreement does track, the disagreement in question is meaningful. The appeal to meaningfulness and coherency in this argument is therefore not meant to presuppose the answer to any philosophical theories of these concepts, but to appeal to a very natural understanding of coherency and meaningfulness that is presumed in all conversation.

P3 is just one such common-sense claim about what does or does not 'track.' I have given examples above demonstrating that distinctively moral disagreements (disagreements over a moral proposition that cannot be explained as a disagreement over some non-moral matter) do track in conversational contexts. Disagreement over a moral proposition when there is no disagreement about empirical observations, the norms of belief formation or the meaning of terms (I have been calling such disagreements distinctively moral) does not cause bafflement in conversational contexts. Disagreement

over an empirically-confirmable proposition under these same circumstances does cause such bafflement. I take the examples of this I discussed above to be sufficient to demonstrate P3.

These are my reasons for endorsing these three premises and, once we accept them, the conclusion that moral propositions cannot be identical to or reduced to empirically-confirmable hypotheses follows. This, in short, is what I take to be the most persuasive argument that moral propositions cannot be identical to or reduced to empirically-confirmable hypotheses. This further implies that moral propositions cannot be confirmed or disconfirmed by empirical evidence alone, as moral considerations are always going to be relevant to their confirmation or disconfirmation.

This conclusion can be accurately understood as a version of the claim that morality is autonomous. Specifically, it is an argument for what has been called the *semantic* autonomy of morality. There are at least three different versions of the claim that morality is autonomous. Charles Pigden describes them as follows:

> For there are, in fact, THREE forms of autonomy: ontological, semantic and logical; and naturalism is only incompatible with the first. Ontological autonomy is the thesis that moral judgments, to be true, must answer to a realm of *sui generis* non-natural properties. Semantic autonomy insists on a realm of *sui generis* non-natural predicates which do not mean the same as any natural counterparts. Logical autonomy maintains that moral conclusions cannot be derived from non-moral premises. (Pigden 1989 p. 128 – emphasis in original)

The argument I have rehearsed above aims to establish only the semantic autonomy of morality. This is all that my thesis requires, as I have only been aiming to establish that propositions that contain moral predicates cannot be identical to or reducible to any empirically-confirmable hypothesis. As I have stated, this argument is neutral on the question of the ontological autonomy of ethics, as it focuses on the nature of non-natural concepts rather than non-natural properties. Furthermore, it is consistent with the denial of the logical autonomy of ethics. (I explicitly denied the logical autonomy of ethics in reference to classical logic in chapter one. However, I concede that there may be some more conservative logical systems, according to which ethics is logically autonomous. The discovery of such a logic would not undermine my arguments to this point.)

It is important to note that this argument in favour of the semantic autonomy of morality diverges quite a bit from Moore's own arguments. Moore's central argument is predicated on the untenable thesis that 'good' is a simple notion that can be easily contrasted with complex notions. But there is a sense in which the argument I have put forward is true to the spirit of Moore's arguments. The cover page of his book displays the telling epigram by Bishop Butler, "everything is what it is and not another thing" (Moore 1903). In relation to Moore, it seems plausible that this is meant to express Moore's contention that many philosophers have mistakenly identified moral properties with natural properties. In my case, however, this quotation takes on a slightly different tone. Rather than emphasizing different types of properties, I have followed Gibbard in emphasizing different types of disagreements. My argument goes further than Gibbard's, as he only establishes that for any proposed definition of a moral term, we can always coherently accept a predication of the definiens while significantly questioning a

predication of the definiendum. This is, in effect, another way of framing Moore's open question argument. The most unique feature of my argument is the claim that the reason moral definitions have a uniquely 'open feel' is that any attempt to define a moral term by using non-moral terms in the context of a moral disagreement mistakes a distinctively moral disagreement for a semantic one. To put my conclusion in another way, I have argued that the possibility of coherent, significant distinctively moral disagreement shows that moral terms cannot be given naturalistic analyses.

Of course, this argument relies on the assumption that moral debate is coherent and meaningful. Some philosophers have suggested that moral disagreement is not cognitively meaningful. This proposal was adopted by early proponents of noncognitivist analyses of moral language. The most famous of these theories was the emotivism proposed by A.J. Ayer (1936) according to which moral judgments are just verbal expressions of emotional states. There are, however, substantive reasons for thinking such a theory is wrong. The most famous of these reasons was proposed by Peter Geach (1965). The so-called 'Frege-Geach' or 'embedding' problem shows that some aspects of our moral discourse cannot be captured by an emotivist theory. Specifically, all deductive inferences that contain moral terms in unasserted contexts (specifically in the antecedents of conditionals) will not be valid if Ayer's emotivism is correct. (See Miller 2003 pp. 40-42 for a detailed discussion of the Frege-Geach problem.) More recent expressivist theories have argued that there is some truth to the emotivist thesis, but they do not go so far as to claim that moral disagreements are not coherent and meaningful (Blackburn 1998, Gibbard 2003). Some philosophers have also contended that moral disagreements are fundamentally confused insofar as they presume the existence of a class of properties

that are not actually instantiated (see for example Mackie 1977). None of the Moorean considerations I have appealed to above are in tension with this claim. The arguments I have canvassed purport to show that moral concepts do not submit to naturalistic analyses. In order to support this thesis, I need to defend only the semantic autonomy of ethics, not the ontological autonomy of ethics. This semantic autonomy is consistent with the kind of error-theory that Mackie proposes insofar as nothing I have said requires that there exist properties that are both instantiated and non-natural. This is the major benefit of following Gibbard and focusing the Moorean arguments on the nature of moral concepts, rather than on moral properties.

In summary, the lesson we can learn from Moorean considerations is two-fold. First, no propositions that have moral content can be analyzed as or reduced to empirical hypotheses. For any amount of empirical evidence that we introduce, we can still meaningfully ask about the moral features of a situation, and this fact about moral discourse is explained by the fact that moral disagreements cannot be reduced to non-moral disagreements. The second point we should take from these Moorean considerations is that there is a long history of moral philosophers failing to recognize the contributions distinctively moral considerations make to moral disagreements. The fact that distinctively moral disagreements cannot be explained as disagreements about meaning, the norms of belief formation, or empirical observations, and thus that moral propositions cannot be reduced to empirical hypotheses, is not always recognized.

It is unfortunate that Moore's own arguments relied on inadequate conceptions of the nature of a definition and the distinction between complex and simple ideas. This fact has led some philosophers to think that Moorean considerations do not establish the

thesis I have argued for. One example of this is Joyce's argument that was discussed above. The argument concludes that the definition of water as H2O remains open in the same way that Moore thinks moral definitions remain open, and thus that the openness Moore identifies cannot prove that there is anything distinctive about our moral concepts. However, my argument does not rely on the assumptions that Joyce criticizes, as my argumentative strategy diverges from Moore's. Rather than relying on Moore's beliefs about the nature of reductive analyses, I rely on claims about the object of discussion in moral disagreements and what does or does not cause bafflement in such disagreements. It is true that an individual with a naïve conception of chemical theory may coherently wonder whether or not water is H2O, but if two people are disagreeing about this identity claim they must either endorse distinct definitions of one of the terms, different norms of belief-formation or they must disagree about some body of empirical observations. The same is not true of individuals who disagree about whether or not moral goodness is identical to, for example, pleasure. There is a meaningful sense in which these individuals can legitimately disagree without us having to conclude that they are talking past each other or that one of them is confused.

Most of what I have argued above should not be extremely controversial. Many philosophers endorse the claim that moral terms do not submit to naturalistic analysis and thus agree that no moral proposition is equivalent to an empirically-confirmable hypothesis. It was, however, important for me to isolate and defend the strongest arguments in favour of this claim. This is because I use the Moorean considerations to defend a much more controversial thesis in the following sections of this chapter. Specifically, I argue that a number of theories about the psychological nature of moral

judgment have moral content and thus cannot be construed as empirically-confirmable hypotheses. The conclusion of these arguments is that theories about the nature of moral judgment are ultimately grounded in moral considerations and that all debate about these questions must be predicated on a recognition of this fact. This implies that moral considerations are the only things that can serve as evidence for or against a theory of moral judgment and thus that empirical science can contribute little to these debates.

I now turn to a discussion of three philosophers who have attempted to inform debates about the nature of moral judgment by appealing to empirical evidence. I argue that none of them succeed in this goal. The theories they discuss can only be confirmed or disconfirmed, if at all, by moral considerations.

IV - Morality, Convention, and Nichols's Sentimental Rules Account

Shaun Nichols has defended a theory of the nature of moral judgment dubbed the sentimental rules account (SRA). His overall goal is to defend the claim that

> …recent evidence on moral judgment indicates that emotional responses do indeed play a key role in everyday moral judgment. However, the emotions themselves are only one part of moral judgment; internally represented rules make an independent contribution to moral judgment. (Nichols 2008 p. 255)

This naturalized sentimentalism is meant to both be "grounded in the empirical evidence" (ibid) and rival alternative metaethical accounts of the nature of moral judgment.

I argue that Nichols's theory cannot be understood as being grounded on empirical evidence. This is because his theory is grounded in certain assumptions that have moral content. The success or failure of his project turns on the plausibility of these moral assumptions. We could therefore reject Nichols's entire project for moral reasons without paying any attention to the empirical research that is supposed to ground it. Insofar as Nichols and other philosophers predicate their metaethical projects on contentious moral propositions, these projects must be understood as being grounded in moral discourse, not psychology or neuroscience.

The main problem with Nichols's work stems from his reliance on the research of Elliot Turiel. Much of Turiel's research is predicated on the assumption that there is a psychologically real distinction between the way that subjects reason about conventional and moral norms. His studies focus on children in various stages of development and observe the differences in how they understand moral and non-moral infractions. The moral domain, Turiel concludes, consists of infractions pertaining to justice, welfare and rights. The conventional domain pertains to the coordination of social interactions, constitutive aspects of social units and social consensus (Turiel 1983 pp. 224-225). He identifies judgments about the justifiability of "killing and the value of life" as moral, but judgments about "modes of greeting (and) forms of address" (ibid pp. 35-36) as conventional. He also asserts that moral judgments make appeals to "concepts of fairness and justice", whereas conventional judgments "pertain to arbitrary acts" (ibid p. 81). More precisely, Turiel defines social conventions as "part of constitutive systems and our shared behaviours (uniformities, rules) whose meanings are defined by the constituted system in which they are embedded" (Turiel, Killen, Helwig 1987 p. 169). Moral rules,

however, are “unconditionally obligatory, generalizable, and impersonal insofar as they stem from concepts of welfare, justice and rights” (ibid pp. 169-170).

These assumptions have moral content insofar as they amount to assertions about the content of the moral domain and, as I will argue below, can be appealed to in order to specify what types of natural properties are morally relevant and morally irrelevant. Because of this, they can be understood as giving moral advice in certain circumstances that would conflict with alternative advice. Furthermore, Turiel explicitly acknowledges that he is borrowing these assumptions from well-known moral philosophers. He justifies this characterization of morality by appealing to the fact that it is “evident in many philosophical treatises” (ibid p. 35). He names Aristotle, Gewirth and Dworkin as adherents of this characterization of moral judgment.

My criticism of Nichols turns on the claim that the assumptions he adopts from Turiel have moral content. The strongest argument in favour of this conclusion stems from the fact there are a number of ethical systems that are in conflict with Turiel’s presuppositions. These systems would, in some circumstances, give moral advice that would conflict with advice that would follow from Turiel’s presuppositions. For example, if one were to argue that judgments about rights, fairness, and justice are not constitutive of moral judgments because some other body of considerations are at the heart of moral thinking, then one could claim that Turiel’s research does not accurately characterize moral judgment. One could, for example, understand certain forms of ethical virtue theory as making such a claim insofar as such a theory could find the locus of moral value in that “which a virtuous agent would do in the circumstances” (Hursthouse 1991 p. 225). While I believe that most versions of virtue theory can make room for

considerations of welfare, justice and rights, these concepts are nonetheless not at the core of moral thought according to virtue ethics. Furthermore, one could articulate a virtue ethic that values strength, independence and courage above all else and disregards considerations of welfare, justice and rights. Such a view is not represented in much contemporary philosophical literature, but it is nonetheless a moral view that would deny Turiel's moral/conventional distinction for moral reasons. According to this view, Turiel is studying some psychologically real class of judgments, but he is mistaken to identify this class as *moral* judgments.

Another viable alternative to Turiel's paradigm would be an ethical system that focuses on the responsibilities that certain relationships give rise to. This view has been suggested by versions of normative care ethics, which claim that approaching moral dilemmas from the perspective of "a network of connection, a web of relationships that is sustained by a process of communication" (Gilligan 1983 p. 32) is a viable alternative to philosophical approaches that assume that moral thought must involve a hierarchical ordering of values, an emphasis on principles or rules, and a presumption of complete impartiality. Like virtue theory, normative care ethics can make room for considerations pertaining to justice, welfare, and rights, but it denies that such considerations are definitive of the moral domain. In many instances, the ethics of care might conflict with an ethical theory that defines morality in terms of impartial, universal considerations of justice and rights. One example of such a conflict would be when my responsibilities to those individuals with whom I am embedded in a network of connection conflicts with my duties to individuals I do not know. Care ethics would emphasize the duties to those individuals with whom I am in relationships, whereas Turiel's framework would

emphasize the importance of impartial consideration for justice and rights. Whereas this conflict may not be incredibly deep between these two interlocutors (they may agree on many points, e.g., both could require that justice be done and both might strongly emphasize the welfare of individuals), this example still serves to demonstrate that normative care ethics denies Turiel's division. Furthermore, such a criticism would be grounded on a rejection of only the moral commitments that serve as a foundation for Turiel's theory. Normative care ethics does not in any way contradict Turiel's empirical evidence, as care ethics does not deny that the psychological phenomenon Turiel identified actually exists. What it *does* deny is that this evidence bears on how we should understand the psychological nature of moral judgment. Whereas Turiel believes that impartiality and universality are essential features of moral thought, care ethics makes room for ethical judgments that prioritize relationships, connection, and communication with concrete individuals. In short, an ethics of care can give us normative moral reasons to reject Turiel's claims about the psychological of moral judgment, without questioning the empirical evidence on which Turiel's theory rests.

Furthermore, Turiel's initial presuppositions are also in conflict with some well-known metaethical theories that I argued above must be understood as having moral content. For example, the claim that there is no distinction between moral judgments and judgments about conventional social agreements would not be without precedent. Some moral relativists have argued that the logic of moral judgments dictates that "the judgment that it is wrong of someone to do something makes sense only in relation to an agreement or understanding" (Harman 1975 p. 3), i.e., a convention. Relativists can reject Turiel's claims about the moral/conventional distinction by stating that the phenomenon

identified by Turiel's empirical studies is one that concerns people's reactions to harm or violations of justice and rights rather than anything about the nature of moral judgment. Once again, we can see how it is that a normative moral theory can accept all of Turiel's empirical findings while denying that these findings reveal anything about the psychological nature of moral judgment. This serves as further evidence for the claim that Turiel's presuppositions have moral content because, as I argued in Chapter Two, it is very difficult to articulate a form of Harman's relativism that does not itself have to be understood as a normative moral thesis.

If we agree on moral grounds that any one of these normative theories is more accurate than Turiel's characterization of the moral domain, then we can reject Turiel's theories about the nature and development of moral judgment. In other words, an adherent of any of these three views could argue that Turiel was not talking about moral judgment after all because he did not accurately identify the moral domain at the outset of his investigation. The only response Turiel could make against such an argument would be a repetition of his claims about what types of things are pertinent to the moral domain and once again citing moral philosophers who agree with him. He therefore cannot fully defend his position without committing himself to substantial claims about the content of moral judgments and thus making contentious moral assumptions about what kinds of things are morally relevant.

None of this was meant to be a substantive criticism of Turiel. His primary goal was to look at the development of moral judgment in children, and to do so he took on a provisional definition of 'morality.' But it is important to note that these assumptions fundamentally ground Turiel's work. Philosophers who predicate their own philosophical

arguments on Turiel's work therefore commit themselves to two distinct theses. The first is that Turiel has identified a real psychological distinction between how people reason about certain domains. The second is that this distinction characterizes the difference between moral and conventional reasoning. The first thesis has come under assault, as Kelly et al. (2007) have performed studies that demonstrate that people's reasoning about the appropriateness of certain acts is not as clearly divided between the two domains that Turiel suggests. I will ignore this controversy. There are, however, some substantial problems with recent work that relies on the second claim, as some philosophers aim to resolve metaethical and moral debates by appealing to Turiel's evidence. One cannot appeal to Turiel's research as a morally-neutral source for empirical evidence about the nature of moral judgment. As I indicated above, one example of this recent work is Shaun Nichols's defense of the sentimental rules account (SRA.)

Turiel's experiments play a major role in Nichols's development of what he considers to be a modest and philosophically neutral account of the capacity for core moral judgment. He explicitly acknowledges his debt to Turiel by stating that "the capacity for moral judgment has perhaps been most directly approached empirically by exploring the basic capacity to distinguish moral violations (e.g., pulling another person's hair) from conventional violations (e.g., chewing gum in class)" (Nichols 2004 p. 5). He also appeals to James Blair's experiments on psychopathy in order to demonstrate that SRA is consistent with these findings. Blair's studies, however, also rely heavily on the moral/conventional task to identify individual subjects' capacity to engage in moral reasoning (Blair 1995, 2005). Nichols' debt to Turiel is therefore quite extensive.

Nichols recognizes that any attempt to draw a distinction between morality and convention would be difficult and contentious. He seems satisfied, however, with saying that these problems constitute "a controversy that we can ignore" (Nichols 2004 p. 5). But by ignoring the moral assumptions built into Turiel's research, Nichols winds up constructing circular arguments against alternative theories, namely, Gibbard's neosentimentalism.

One central component of Nichols's defense of SRA is the rejection of a theory which I have argued must be understood as having moral content: Gibbard's neosentimentalism. This theory interprets moral judgments as judgments about the appropriateness of feeling certain emotions towards particular acts or state of affairs. As Gibbard puts it, "what a person does is *morally wrong* if and only if it is rational for him to feel guilty for doing it and for others to resent him for doing it" (Gibbard 1990 p. 47). Nichols's argument against this form of neosentimentalism relies on the empirical evidence that very young children can pass Turiel's moral/conventional task, and thus are capable of making core moral judgments prior to the development of an understanding of guilt (Nichols 2003 p. 90).

Nichols is wrong to think that the disagreement between him and the neosentimentalist can be resolved by an appeal to empirical data. Insofar as Turiel's work, in particular his assumptions about the moral/conventional distinction, relies on a prior articulation of what the content of the moral domain is, Nichols's argument can be rejected by anyone who denies that the only things that are morally salient are harm, welfare, justice, rights and fairness. Insofar as neosentimentalists hold that moral judgments are those judgments directed at acts toward which it is rational to feel guilt,

proponents of neosentimentalism would deny that the psychological distinction identified by Turiel has anything to do with the distinction between moral and non-moral judgments. The psychological distinction Turiel identified might be, as Gibbard elsewhere suggests, that between norms that rely on authority and those that do not (Gibbard 2006b p. 202), not the distinction between the moral and the non-moral. According to Gibbard, children who lack an understanding of guilt really would lack an ability to make core moral judgments.

What, then, is really at issue between the neosentimentalist and the proponent of SRA? It is clearly not a matter of empirical fact, as both recognize that children draw the distinction Turiel indicated. It is, I propose, a moral disagreement. Insofar as the neosentimentalist identifies moral wrongness with those things towards which it is appropriate to feel guilt or resentment, she disagrees with the proponent of Turiel's theories. Recall that Turiel identifies moral judgments as those which do not have an 'arbitrary' act as their object. According to Turiel, any act that is not arbitrary therefore has to do with justice, rights or welfare. Any judgment with this kind of content is moral according to Turiel's paradigm. But this is precisely what the neosentimentalists want to reject insofar as they claim that considerations regarding justice, rights and welfare can come apart from moral considerations. Insofar as Nichols relies on experimental data that presuppose certain moral theses the neosentimentalist would reject, he begs the question against the neosentimentalist. Nichols does so because he does not recognize that Turiel's work presupposes substantive moral content and thus cannot be construed as an empirical account of moral judgment.

To summarize, any philosophical theory of moral judgment that relies on Turiel's evidence must acknowledge that his account of the nature of moral judgment is grounded in normative ethics. Such a theory could therefore be rejected if those moral foundations conflict with considered moral judgments about what belongs in the moral domain. Furthermore, it is extremely difficult to apply empirical evidence to the debate about the nature of moral judgment without assuming a number of contentious moral propositions. Because of this, Turiel's evidence cannot be appealed to in order to support a philosophically interesting account of moral judgment without begging the question against opposing philosophical theories. Nichols's SRA is therefore not fundamentally grounded primarily on empirical evidence, but rather on substantive moral assumptions.

I now turn to Prinz's empirically-informed defense of a sentimentalist moral relativism. Like Nichols, Prinz grounds his theory on moral assumptions. In addition to this, Prinz's theory about the nature of moral judgment itself has moral content and he acknowledges this by claiming that, contrary to the arguments I raised above, moral propositions can be confirmed by empirical evidence. For these reasons, my primary disagreement with Prinz is more fundamental than my disagreement with Nichols.

V - Prinz's Sentimentalist Relativism

It is plausible to attribute to Prinz the belief that questions about the nature of our moral beliefs, judgments, and practices are entirely empirical in nature and can be confirmed or disconfirmed only by empirical evidence. For example, he predicates his metaethical theory on the following assumption: "figuring out what we believe about

morality is a descriptive task *par excellence*, and one that can be fruitfully pursued empirically" (Prinz 2007a p. 1). In order to accomplish the goal of describing our moral beliefs, judgments, and practices, Prinz endorses a methodological stance he refers to as methodological naturalism, which assumes that all facts (including those about morality) are natural facts, and thus requires that "the methodology by which we investigate facts must be suitable to the investigation of natural facts" (Prinz 2007a p. 3). This version of methodological naturalism is taken by Prinz to imply that we "should investigate norms using all available empirical tools" (ibid). He explicitly endorses this methodological principle when he states the following: "my most obvious commitment is to methodological naturalism, because I will draw on empirical findings…including findings from neuorscience, psychology, psychiatry, anthropology, cultural history, and ethology" (Prinz 2007a p. 9). He regularly appeals to such empirical evidence in defense of his own theory.

The theory he settles upon is committed to a form of sentimentalism that combines two sub-theses about the role that emotions play in morality. The "metaphysical thesis" is an identification of moral rightness with those properties that cause "feelings of approbation in normal observers under certain conditions" (Prinz 2007a p. 20). The "epistemic thesis" states that the disposition to feel the emotions described in the metaphysical thesis "is a possession condition on the normal concept RIGHT" (ibid p. 21). The conjunction of these two views is constitutive of a metaethical theory that Prinz endorses and dubs 'strong emotionism.' He argues for this view by appealing to a large amount of interdisciplinary empirical research that he takes to collectively demonstrate that moral judgments are accompanied by emotions (ibid pp.21-

23) and that levels of emotional affectivity exert influence over moral judgments, even when the emotional arousal was caused by morally irrelevant factors (ibid pp. 23-29). He then appeals to other psychological literature from the study of the moral development of children (ibid pp. 32-37) and psychopathology (ibid pp. 42-47). Many of these discussions are predicated upon studies that presume the correctness of Turiel's distinction between morality and convention (see, for example, Prinz's discussion of Turiel (1983) on pp. 35-36 and his discussion of Blair (1995) and Blair et al. (1997) on pp. 44-45). The problems with reliance upon Turiel's paradigm in order to justify a philosophically illuminating account of moral judgment were discussed above, so I shall not further address them here. It is, however, important to note that these considerations play a large role in Prinz's summary of the evidence in favour of strong emotionism and thus that Prinz's entire theory is predicated on certain morally-loaded assumptions. This point will play a role in the arguments to come.

Prinz then goes on to argue in favour of what he considers to be a version of sensibility theory like that put forward by John McDowell (1985). The theory he proposes revolves around the claim that having a moral sentiment against X in one's long-term memory is constitutive of having internalized a moral rule against X. Moral properties, according to this theory, are defined in terms of the emotional dispositions of observers. More rigorously, Prinz proposes the following formulation of his theory:

> An action has the property of being morally wrong (right) just in case there is an observer who has a sentiment of disapprobation (approbation) towards it. (Prinz 2007a p. 92)

Moral judgments are, according to Prinz, just manifestations of particular emotional dispositions (ibid p. 96). He thinks that this thesis, which he takes to follow from the proposal above, is empirically-confirmable. As support for the proposal, he identifies a plethora of empirical considerations that he believes lend credence to it (ibid p. 97).

However, Prinz's proposal clearly has moral content. It identifies the truth conditions of many moral propositions, picks out a set of naturalistic properties as being uniquely morally salient, and it would provide moral advice in many situations that would be in tension with the advice given by a number of other moral theories. Because of this, it is not clear what role empirical evidence can play in the vindication or refutation of this view.

Prinz's case is different from Nichols' for two reasons. First, Prinz's theory is not only predicated upon moral assumptions, but also itself has moral content. Nichols's SRA is sufficiently vague that I believe it could be understood as lacking moral content. Prinz's relativistic proposal, however, can clearly be used to give moral advice in cases wherein observers have sentiments of approbation or disapprobation and this advice could conflict with that given by alternative moral theories.

Second, my criticism of Nichols was that his theoretical presuppositions carried moral commitments that he did not recognize. Prinz, however, explicitly acknowledges that his relativistic theory has moral content, but he does not believe that this feature of the analysis disqualifies it from being considered an empirically-confirmable hypothesis. He says, for example, that whenever we correctly identify a sentiment in ourselves "we

are discovering wrongs empirically" (Prinz 2007b p. 284). He also expresses this view when he articulates his philosophical ambitions in the following manner:

> I will argue that morality derives from us. The good is that which we regard as good. The obligatory is that which we regard as obligatory. The 'we' here refers to the person making a moral claim and the cultural group with which that individual affiliates. If the good is that which we regard as good, then we can figure out what our obligations are by figuring what our moral beliefs commit us to. Figuring out what we believe about morality is a descriptive task *par excellence,* and one that can be fruitfully pursued empirically. Thus, normative ethics can be approached as a social science. (Prinz 2007a p. 1)

This conclusion is clearly in tension with the arguments I raised above regarding the distinction between moral propositions and empirically-confirmable hypotheses. In contrast to my conclusion that empirical considerations cannot bear on moral matters, Prinz concludes that because his theory is both empirical and moral, empirical considerations can and do bear on moral matters.

My disagreement with Prinz can be summarized in the following manner. Prinz believes that disputes over propositions that have moral content can only be resolved by appealing to empirical evidence. Or, as Prinz puts it: "here, as elsewhere, we are hostage to empirical fortune, and I think relativism is supported by the evidence" (Prinz 2007b p. 285). Admittedly, the evidence I have given above only suggests that Prinz believes that empirical data can make *some* contribution to these debates, not that empirical considerations are the only kind of consideration that matters. However, at other times, Prinz does seem to commit himself to the claim that *only* empirical data is relevant to the

adjudication of these theories. He agrees, for example, that conceptual analysis plays some role in the defense of metaethical and normative moral theories, but he also believes that this traditional philosophical tool acquires legitimacy only from the fact that it approximates an empirical mode of inquiry: “conceptual analysis is, like all legitimate investigatory tools, an empirical method” (Prinz 2007a p. 3). However, Prinz also disparages conceptual analysis by claiming that “as empirical methods go, it is not especially powerful” (ibid). According to Prinz, then, all metaethical and moral investigation, if it is to be legitimately conducted, must be investigated empirically. Furthermore, Prinz thinks that empirical evidence always trumps our considered moral judgments and intuitions about ethics: “we do not like the idea that people are obligated to do things that we find morally repellent. If the theory has that consequence, we will regard it as very unfortunate, but that does not make the theory wrong” (Prinz 2007b p. 285). In short, it is plausible to attribute to Prinz the belief that moral and metaethical theories can be confirmed or disconfirmed only by empirical evidence.

Alternatively, I claim that an understanding of moral discourse that insulates morality from empirical considerations and upholds the autonomy of ethics is preferable insofar as such an account allows us to make sense of the practicality of moral judgment and the significance of distinctively moral disagreement. According to my view, the fact that Prinz’s theory of the nature of moral judgment entails so many substantive moral propositions implies that it cannot be the sort of proposition that can be confirmed solely by appeals to empirical evidence. These claims rely strongly on the arguments I articulated above, which I will not rehearse here. Should we side with Prinz on this

matter, or should we endorse my account? I think there is good reason to prefer my account over Prinz's.

The major reason why I think we should, contra Prinz, endorse the claim that moral propositions cannot be identical to or reducible to empirically-confirmable hypotheses is that I believe the 'What's at issue?' argument to be successful, and it implies that there is a rigid distinction between moral and naturalistic propositions. Prinz does not and, I will argue below, cannot offer a response to the Moorean argument I articulated above. The closest he comes to addressing these concerns is offering an argument that we can derive some propositions containing the word 'ought' from premises that do not contain the word 'ought' (Prinz 2007a p. 5). But this claim does not relevantly address the Moorean considerations. In Chapter One, I acknowledge that there are some arguments with non-moral premises that nonetheless have moral conclusions. As I stated above, the Moorean argument I offered above does not rely on any claim to the contrary (it is only meant to establish semantic, not logical, autonomy.) Prinz's argument, then, does not establish that the central force of the 'What's at issue?' argument should be rejected. If we accept that argument, we should reject Prinz's thesis.

There is, however, another potential response to my Moorean argument that Prinz could put forward. His response to the 'What's at issue?' argument is implicit in his response to a related problem: he is committed to a number of conclusions that seem to run counter to many plausible beliefs about moral discourse. His thesis entails that a number of thoughts and beliefs that seem quite natural and coherent are in fact incoherent. For example, under Prinz's interpretation, any individual who asks himself 'I know I think X is morally wrong, but am I mistaken?' has asked an incoherent question:

> A member of the Yedizi can say, ‘I know I morally value honor killing but should I? Is it really something I am obligated to do?’ It is a consequence of my theory that this kind of question is really incoherent. That may look like a fatal objection to the theory, but it is not an embarrassment if there is a reasonable explanation of why such questions seem intelligible to the people who formulate them. Fortunately, a reasonable explanation is available. Recall that I am not committed to the view that people realize morality is a response dependent property. People can (and often do) believe that morality is mind-independent. (Prinz 2007b p. 288)

In short, Prinz’s response to this kind of concern is to endorse an error-theoretic stance towards common-sense moral discourse. This position implies that all sincere moral debate rests on a mistake. When we argue with someone about moral matters and thus engage in distinctively moral disagreement, we are either using mistaken concepts that presume the mind-independence of morality or we are incorrectly assuming that our interlocutor has sentiments that are the same as our own. This response does not directly address the Moorean argument I offered above as that argument does not explicitly presume the mind-independence of morality, but Prinz’s response to my Moorean argument can be glossed from the paragraph quoted above. If pressed on my Moorean argument, Prinz would simply claim that the philosophers I have named, and anyone else who is persuaded by the Moorean considerations, have false beliefs about morality in just the way described above. The difference is that their mistake is not the assumption that morality is mind-independent, but that distinctively moral disagreement (moral disagreement that cannot be reduced to non-moral disagreement) is coherent. Should we be persuaded by this response? I argue below that we should not.

Prinz’s error-theoretic response is not sufficient for refuting the Moorean argument I articulated above because it presumes that his own account of the nature of moral judgment is correct. This, in and of itself, does not allow us to reject Prinz’s theory, as it would be just as unfair for me to assume at the outset that his account is not true as it would be for him to assume that it is. The criticism I wish to raise against Prinz pertains to how he goes about arguing in favour of his account of moral judgment. As I demonstrated above, Prinz understands his response-dependent account of moral judgment as being confirmed by empirical evidence and he believes, as he puts it, that empirical evidence holds philosophers hostage. As we saw above, Prinz considers all legitimate inquiry to be empirical in nature. Since Prinz takes his theory to be supported by the strongest empirical evidence available, he therefore treats his theory as being on better epistemic footing than philosophical theories that are not supported by empirical evidence (such as theories that maintain that common-sense moral discourse is not fraught with error and endorse one of the Moorean arguments discussed above.) However, upon closer inspection of Prinz’s premises (which will occur below), we find that he begins his argument by assuming the truth of certain propositions that turn out to have moral content. Once we recognize that Prinz’s entire theory is grounded on some controversial intuitions that cannot be understood as being confirmed by empirical evidence, this fact undermines his claim that his own theory has better epistemic footing than alternative theories. In fact, since the assumptions that support the Moorean arguments discussed above are far more intuitively plausible than Prinz’s proposal, we actually have more reason to accept them than we do Prinz’s assumptions. I now turn to

the evidence that Prinz's argument depends heavily on moral intuitions that are not, and cannot be, supported by empirical evidence.

As I noted above, Prinz relies heavily on Turiel's work. Turiel's studies, however, are predicated upon assumptions that have substantive moral content. Appealing to Turiel's work is dangerous for philosophers insofar as any account predicated on Turiel's research must accept a number of controversial moral propositions. If philosophers reject those theses, they can similarly reject any conclusions that are predicated on acceptance of Turiel's moral/conventional distinction. But Prinz does not rely exclusively on Turiel. He also aims to support his emotionism by appealing to a number of studies that purportedly show that "emotions can influence moral evaluations even when the emotions are induced by morally irrelevant factors" (Prinz 2007a p. 28). The problem with this type of approach is comparable to the problem associated with relying on Turiel's research: we cannot apply these findings to moral judgment without first endorsing a particular account of what types of things count as 'morally irrelevant factors.' Such an account must therefore presume that certain types of factors are morally relevant and is therefore committed to an account of the substance of morality. This presupposition, insofar as it identifies certain features of the world as being uniquely morally salient and rules out alternative moral theories, must be understood as having moral content.

Prinz's entire account is therefore predicated on some premises that must be understood as substantial moral theses. If we reject these premises we have no reason to admit that the empirical considerations Prinz appeals to confirm his theory. In other words, rather than thinking that moral philosophers are held hostage to empirical

findings, it seems that Prinz's project is held hostage to some substantive moral beliefs. Furthermore, Prinz cannot conclude that those assumptions are justified by empirical evidence because he appeals to these assumptions in order to apply the empirical findings to the metaethical questions at hand. Prinz is therefore wrong to state that the counterintuitive moral implications of his theory do not imply that his theory is wrong, as the theory itself is predicated on the acceptance of moral premises that can be legitimately rejected if they are found to be substantially counterintuitive (for example, if they are found to contradict true moral propositions).

Since Prinz cannot claim that his theory follows from a body of empirical evidence, he also cannot claim that such a body of evidence invalidates the intuitions of philosophers who find the Moorean considerations persuasive. Contrary to his own belief, Prinz is not on better epistemic footing than his philosophical opponents. If Prinz is justified in assuming a number of contentious moral theses at the outset of his project, then he cannot consistently claim that his opponents are not justified in doing the same. It would be arbitrary for him to conclude that his initial moral, intuitive assumptions are legitimate but to claim that the intuitions about the intelligibility of moral disagreement that ground the Moorean considerations are illegitimate. His argument that this presupposition is grounded on a mistake is really no argument at all, but is rather the assumption that his presuppositions about the nature of moral judgment take priority over the presuppositions of other philosophers. Further, since his view is committed to denying that many meaningful, significant questions about the status of moral disagreements are incoherent, alternative theories that can account for the meaningful

nature of these moral disagreements are actually on better footing than Prinz's own theory.

Prinz is therefore wrong to treat his theory as an empirically-confirmable hypothesis. Prinz cannot account for the force of the Moorean considerations outlined above, his theory has some profoundly counterintuitive implications and his entire account turns on a number of contentious moral propositions which are not themselves empirically-confirmable. Prinz's entire project can be rejected if we reject his moral starting points. These considerations collectively support my conclusion that the appropriate way to resolve the disagreement between myself and Prinz is to assess our starting assumptions on the basis of their intuitive plausibility, since nothing else can validate or invalidate these assumptions. Empirical evidence does not provide much help in this case, and to pretend that it does is to fail to recognize the non-empirical, intuitive assumptions that actually shape our views about these matters.

I take these arguments to sufficiently demonstrate that we should endorse my view that no moral propositions are empirically-confirmable and reject Prinz's claim that his theory has a superior epistemic footing because it can be construed as an empirical hypothesis.

As one final response to this argument, a proponent of Prinz's view might claim that the error-theoretic thesis discussed above actually makes Prinz's argument look more revisionist than I have here acknowledged. In other words, it is possible that Prinz's aim is not to adequately describe morality, but rather to advise that we stop using our moral concepts and offer a prescriptive account of what concepts we *should* use instead. But

this approach not only contradicts Prinz's stated goals (Prinz 2007a, p. 8), it also amounts to a concession that my argument above is correct. Recall once again that my Moorean argument does not presume that moral properties are instantiated in the world. It is simply a claim about how our moral concepts work, or, in other words what does or does not track in conversational contexts. If we take Prinz to be suggesting that this Moorean analysis is a correct account of how our moral concepts work, but that we ought to abandon these concepts, then he has essentially conceded that my analysis is correct. My thesis is neutral on the question of whether or not we should persist in making moral judgments, so a revisionist thesis such as this does not refute the arguments I have made above.

There is, however, an entirely different way that proponents of Prinz's view might go about arguing against my position. Some philosophers have argued that Prinz and I are both incorrect in thinking that Prinz's version of relativism has moral implications. Valerie Tiberius, for example, has argued that we are wrong to think "the kind of relativism we get from Prinz's metaethics matters to normative ethics and moral practice" (Tiberius 2009 p. 723). If Tiberius is right and Prinz's position has no bearing on ethics and morality, then the arguments above against Prinz's theory are predicated on a false presupposition. I close my discussion of Prinz by considering the implications of Tiberius's thesis for the arguments I have articulated above.

The major assumption at the heart of Tiberius's argument is that the justification of a moral standard can be completely independent from its truth conditions. If this is true, then Prinz's theory can be understood as specifying only the truth conditions of moral propositions without saying anything about what justifies a moral standard. She

grounds this assumption by claiming that it follows from Prinz's discussion of the possibility of moral progress. In order to understand the force of her argument, we must first discuss some details of the mechanics of Prinz's system.

According to Prinz's account, the truth conditions of moral propositions are relative to individuals. The obvious problem with this is that it then seems impossible to objectively say of any moral system that it is better than any other. This threatens the possibility of understanding historical moral developments as progressive (as opposed to being mere changes), as it seems as though Prinz's theory implies that these developments are mere changes. However, Prinz still wants to allow for the possibility of moral progress and the idea that some moral systems are better than others.

Prinz's commitment to the existence of moral progress is initially problematic for him, as he believes that we cannot legitimately say of individuals who disagree with us about moral matters that they morally ought to do anything. This feature of Prinz's view is important for Tiberius's argument. Prinz explains the difference between judgments about moral wrongness and moral 'ought' judgments in the following way:

> A judgment that X *ought* to ϕ is true if and only if it is wrong not to ϕ on the value systems of both the speaker and X
>
> A judgment that ϕ-ing is wrong is true if and only if the ϕ-ing is the object of a sentiment of disapprobation among the contextually salient individual(s) (usually the speaker) (Prinz 2007a p. 180)

Both formulations are relativistic, but they are distinct in an important way, namely, ought judgments do something more than wrong judgments. Judgments of wrongness,

according to Prinz, are nothing more than judgments about what sentiments the speaker actually has (see above formulation), whereas "an ought judgment conveys the fact that a norm has authority over the behaviour of the person addressed by that judgment." (ibid p. 178). Ought judgments therefore presuppose that an individual other than the judger accepts a certain norm, whereas wrongness judgments do not.[37]

It seems fairly obvious that this analysis leads to the counter-intuitive conclusion that we can never say of any particular moral system that it is better than another in any robust sense insofar as we cannot say of anyone who disagrees with us that they morally ought to do anything. Prinz's solution to this is to recommend that we "subject our values, including those we treasure, to rigorous reconsideration in light of extramoral concerns." (ibid p. 302) Some examples of extramoral concerns are our wants and needs (ibid p.307). One specific example Prinz relies on is the kinds of lives that we find desirable (ibid p. 299). Other examples are "consistency, coherence with facts, stability, ease of implementation, welfare, well-being, completeness, universality, genealogical impunity, and conformity to biological norms" (ibid p. 292). Tiberius appeals to these and other considerations to conclude that "even if the truth of a moral judgment is relative to the speaker, it is still the case that our judgments can be better or worse and that we can revise them to conform to the standards for better" (Tiberius 2009 pp. 723-

[37] To my own ear, this account is not persuasive. I understand judgments that one morally ought to do X as equivalent to judgments that it would be morally wrong to do anything but X. Prinz would respond to this by once again asserting his relativistic analysis and claiming that the empirical evidence directs us to it. But there are two problems with this appeal. First, Prinz's account is not purely empirical as it begins by assuming substantive moral premises. Second, these claims about when one may conclude that an individual morally ought to do something seem to fall under the class of moral propositions. This means that they can be rejected for moral reasons. Furthermore, they are counterintuitive moral propositions that do not fit with my own, and I suspect many other peoples', reflective moral beliefs and thus they can be legitimately rejected. Contra Prinz, I believe that all individuals ought not to infibulate their children regardless of whether or not the individuals in question agree with me, and I would extend this consideration to many other moral agents with defective moral beliefs.

724). In other words, there are extramoral considerations that can be appealed to in order to determine whether or not one moral system is better than others. This process of justification, which is supposed to be distinct from a quest for moral truth, is what allegedly justifies the claim that Prinz's relativism is irrelevant to moral practice.

According to Tiberius, regardless of whether or not a particular moral standard X is true in a given context, it can still be justified. This would be the case when the following two conditions obtain:

1) If a moral standard X satisfies extramoral considerations A, B and C, X is justified.
2) Moral standard X satisfies A, B and C.

Because of this, Tiberius holds that Prinz's claim about the truth conditions of moral propositions has no bearing on moral practice, since the justification-conditions and truth-conditions of moral standards come apart. The problem for this argument is that there is a very clear sense in which a proposition that has the form of (1) must be understood as giving moral advice. If at some point I am unsure of which moral standard I should appeal to in a situation, I can look to (1) to tell me what features of the standard should be relevant to my decision. Any choice of how to fill in the variables 'A', 'B' and 'C' in (1) will itself have moral content. That means any judgments of the form (1) will be moral judgments and, if Prinz is correct, will therefore have to be understood according to his relativistic model.

A consequence of the fact that any instantiation of (1) would have moral content is that Prinz is skewered on the horns of a dilemma. His first option is to acknowledge that even judgments about the justification of moral standards must be understood as internal to moral discourse and thus must be analyzed relativistically. Such an acknowledgment would imply that the justification of moral standards is dependent upon which sentiments we have regarding that standard. If this is the case, then one can appeal to Prinz's theory about the nature of moral judgment in order to determine in what circumstances an instantiation of (1) is correct (i.e., just in case there is an observer who has a sentiment of approbation towards it). If Prinz embraces this horn of the dilemma, then his theory about the nature of moral judgment would have moral content, and thus my argument that Prinz's theory is not an empirically-confirmable hypothesis would go through.

Alternatively, Prinz could deny that judgments about the justifiability of moral standards succumb to relativistic analyses. We have, however, already established that such justificatory judgments must be understood as being internal to moral discourse. This would, in effect, amount to Prinz rejecting his own analysis insofar as he would have to admit that his theory does not capture the form and content of some moral judgments. In short, Tiberius's claim that Prinz's relativism has no implications for morality does not protect Prinz from the criticisms I have here raised.

VI - Joyce's Minimal Projectivism

In addition to Prinz and Nichols's attempts to apply the findings of empirical moral psychology to metaethical debates, some other philosophers have attempted to refine philosophical theories about the nature of moral judgment to the point where they can be directly tested as psychological hypotheses. This is the approach Joyce has taken towards a version of moral projectivism. I argue below that he also conflates some moral propositions with empirical hypotheses. Because of this mistake on Joyce's part, his project fails.

Moral projectivism is a theory committed to explaining why moral properties seem objective. Since the terms 'objective' and 'subjective' are used in many different ways, it will be useful for me to make explicit the sense of 'objectivity' Joyce uses. I follow Joyce in using a relative notion of objectivity that emphasizes the subject-independence of objective properties. A property X is objective relative to an individual Y when Y understands X as a property that exists in the world and "is there to be discovered, of which (Y) might be ignorant, of which (Y) is a passive observer (and) which *could* have occurred without (Y)" (Joyce 2009 p. 64). Conversely, a property X is subjective relative to an individual Y when these things are not true of X for Y. A property might be dependent on a subject (such as the psychological property of sadness) and yet be understood as an objective property relative to some observer (such as when one discovers the sadness of another). This relative conception of objectivity and subjectivity should be sharply contrasted from what Joyce refers to as 'absolute' theories of objectivity and subjectivity according to which properties can be understood as

objective or subjective from all standpoints. Since all subjective phenomena are objective relative to some other standpoint, Joyce believes that an absolute conception is not useful in helping us understand the phenomena discussed by projectivists:

> After all, what we intuitively want to capture of the objectivist tendency is the experience a person may have that 'I am not the author of this phenomenon; it could carry on the same even unperceived by me.' We do not have to worry about the absolute 'objective' status of the phenomenon (or, indeed, whether it is even coherent to think of any such notion of absolute objective status standing in contrast to absolute subjective status); we need concern ourselves only with how the subject experiences it in relation to herself. (ibid p. 64)

According to this understanding, the objective/subjective status of a property relative to an individual is determined by the character of the experiences the individual has of that property. My own sadness is subjective relative to me because of the character of my experience of my own sadness and it is objective to others because of the character of their experiences of it.

Joyce believes that the central commitments of any projectivist theory can be understood as a conjunction of empirical hypotheses about human psychology. The two psychological hypotheses he isolates are:

> The Phenomenological Thesis: We experience moral wrongness (e.g.) as an objective (in the sense explained) feature of the world. [38]

[38] Treating the phenomenological thesis as a psychological hypothesis is a common theme in the philosophical literature. Arguments that are premised on this assumption are so widespread that several philosophers have called this methodology into question (See, for example, Loeb (2007) and Kirchin (2003)). Psychological readings of the phenomenological thesis have been used to support many different

> The Causal Thesis: This experience has its origin in some non-perceptual faculty; in particular, upon observing certain actions (etc.) we have an affective attitude (e.g., the emotion of disapproval) that brings about the experience described in the phenomenological thesis. (Joyce 2009 p. 56)

Joyce refers to the conjunction of these two theses as *minimal projectivism*. He argues that the phenomenological thesis is an empirically-confirmable hypothesis.

Joyce does an admirable job of dealing with the concerns "that there is something vague, vacuous, ambiguous or incoherent about the (phenomenological) thesis." (ibid p. 68) I believe that he successfully addresses the ambiguities associated with the terms 'objectivity' and 'experience',[39] but there is a further problem with his claim that the phenomenological thesis is empirically-confirmable. The problem in question is that the phenomenological thesis has moral content.

It may seem intuitively unlikely that the phenomenological thesis is best interpreted as having moral content. It is far more obvious that Turiel's presuppositions and Prinz's theory are moral, as there are clearly cases in which they can be used to offer

---

positions by both realists and anti-realists. The thesis is generally interpreted in one of two ways. The first interpretation understands the phenomenological thesis as a claim about *what it is like* to engage in moral deliberation or to make moral judgments. Jonathan Dancy, for example, explains the phenomenological thesis as being about "the nature of experience." (Dancy 1986 p. 172) David McNaughton similarly asserts that "our experience of the world does seem to involve experience of value" (McNaughton 1988 p. 19). For both of these philosophers, the phenomenological thesis is interpreted as an assertion about the nature of the phenomenal consciousness that accompanies moral cognition. The second interpretation does not understand the thesis as addressing phenomenal consciousness, but as being about how we engage in moral deliberation and speak about moral beliefs. David Brink, for example, writes that we often "deliberate as if there is a correct answer to the (moral) question before us" (Brink 1989 p. 36) and claims that our talk of moral knowledge tends to presuppose a form of moral realism according to which certain things are objectively right or wrong. J.L. Mackie also premised the argument for his error theory on the presupposition that the phenomenological thesis has a "firm basis in ordinary thought, and even in the meaning of moral terms" (Mackie 1977 p. 31).

[39] See Joyce 2009 pp 62-68.

moral advice that would be in tension with alternative moral advice. At first glance, the phenomenological thesis does not seem to be offering moral advice. But on closer investigation, it turns out that certain forms of the moral standard commonly referred to as speaker subjectivism offer moral advice that conflicts with the phenomenological thesis. In the following paragraphs, I will describe one case in which believing both the phenomenological thesis and a form of normative moral subjectivism would lead one to make contradictory moral judgments. I will then discuss why, in light of this result, it makes sense to think of the phenomenological thesis as having moral content.

By ‘subjectivism’ I mean a moral standard that identifies a moral property with a subjective property in Joyce’s sense of the term. Consider Joyce’s example of a subjective property, one’s own sadness. This is a subjective property insofar as I do not experience my own sadness as an objective feature of the world. From my own standpoint, it is clear that my sadness depends on me and could not carry on without me. It is these features that make my own sadness subjective relative to me. In any world where I do not experience sadness, the property ‘my sadness’ has changed or ceased to exist. Similarly, if moral properties are subjective properties, we would have to agree that any world wherein the relevant mental state changes, the moral property must also change. A subjectivist moral theory in Joyce’s sense of the term “subjective” would have to hold that a moral property is constituted by an individual’s “act of apprehending it” (ibid p. 63). If such an act did not occur, the relevant moral property could not exist.

The phenomenological thesis contradicts at least one such subjectivist moral theory. This theory states the following:

> SUBJECTIVISM: The property of being morally impermissible is identical to the property of being the object of my disgust.

The disgust mentioned in SUBJECTIVISM is understood subjectively. As a moral standard, SUBJECTIVISM effectively recommends that one appeal only to one's own disgust in order to determine what is morally relevant. Further, SUBJECTIVISM states that successful moral deliberation is accomplished by paying attention to one's own affective attitudes. Disgust, as it is understood in SUBJECTIVISM, is the sort of thing that is not experienced as an objective feature of the world. When I identify something as the object of my own disgust, I do not experience that property as an objective feature of the world, but rather as a subjective property. If SUBJECTIVISM were to be true, we would experience moral impermissibility in the same way that we experience our own disgust. If this were the case, the phenomenological thesis would be wrong about the nature of moral phenomenology.

Furthermore, SUBJECTIVISM is a proposition with moral content because it provides moral advice that is tension with the moral advice that would be offered by alternative moral theses. For example, it contradicts the following universal hedonistic moral principle:

> HEDONISM: All and only acts that maximize pleasure in the universe are morally permissible.

HEDONISM and SUBJECTIVISM would ascribe different moral predicates to identical situations. In a situation where I am disgusted by an act that would maximize pleasure, SUBJECTIVISM and HEDONISM would generate contradictory moral judgments. Both SUBJECTIVISM and HEDONISM therefore make a claim about what sort of property moral impermissibility is. Since HEDONISM understands the property of moral permissibility as being identical to the property of being one of those things that maximize pleasure, it treats moral properties as objective in Joyce's sense. From any individual's perspective, the pleasure of others is there to be discovered, can occur without that individual, and so on. SUBJECTIVISM, however, understands moral properties as subjective properties.

The fact that SUBJECTIVISM has moral content is important because one cannot believe the phenomenological thesis and SUBJECTIVISM at the same time without applying contradictory moral predicates to a single situation. For example, I might turn the corner one day and see a group of individuals performing a harmless act. When I perceive this act, I do not have an experience of an objective property of wrongness. Despite this, I find myself having uncharacteristic feelings of disgust towards the act in question. If I endorse the phenomenological thesis, I must understand the objective-seeming properties as the only types of properties that are relevant to my making a moral judgment. If I endorse SUBJECTIVISM, I must understand my own disgust as the moral arbiter in this case. The two propositions serve as contradictory moral standards insofar as they offer conflicting moral advice in this situation.

According to the phenomenological thesis, the only things that could count as moral impermissibility are those that I experience as an objective feature of the world. Since in the case described above I do not experience any objective-seeming property of

moral wrongness, but rather only experience subjective-seeming moral properties, I must conclude that the act is morally permissible (assuming, of course, that any act which is not wrong is permissible). If I endorse SUBJECTIVISM, however, the only thing that could count as a moral property is my feeling of disgust. SUBJECTIVISM requires that in this circumstance, I judge the act to be morally impermissible. Since the phenomenological thesis and SUBJECTIVISM yield different moral pronouncements in this case and SUBJECTIVISM has moral content, the phenomenological thesis must also have moral content. Otherwise such a disagreement between the two standards would not occur.

Another way of putting this argument is by pointing out that SUBJECTIVISM and the phenomenological thesis identify different classes of properties as being uniquely morally salient. While SUBJECTIVISM implies that moral impermissibility is the kind of thing that is experienced as a subjective property (I experience moral impermissibility in the same way that I experience my own disgust), the phenomenological thesis implies that moral impermissibility is the kind of thing that is experienced as an objective property. The phenomenological thesis therefore implies that the property of moral wrongness is not identical to the property of being the object of my disgust, and therefore rules out a number of possible normative moral theories which hold that these properties *are* identical. This is sufficient for concluding that it has moral content.

To summarize the argument, the phenomenological thesis implies that only things we experience as objective features of the world are relevant to moral deliberation, as it characterizes our experience of moral wrongness as being about objective features of the world. However, some moral standards are committed to the claim that some things that we do not experience as objective features of the world, such as one's own disgust, are

uniquely relevant to moral deliberation. The phenomenological thesis can therefore be understood as giving moral advice that conflicts with the advice given by SUBJECTIVISM.

One plausible objection to the claim that the phenomenological thesis has moral content is that it is still possible that moral properties are subjective properties which we experience as objective properties. The fact that we experience something in a certain way does not, after all, entail that it *is* that way. When we look at an oar that has been partially submerged, we have the experience of seeing a bent oar. But the fact that we experience the oar as bent says nothing about what it is actually like. The same may be true for moral properties. We might experience them as seeming objective despite the fact that they are subjective. This, after all, seems to be the point of the projectivist thesis. It may be possible, in other words, that our own disgust becomes 'projected' on to the world and is then, despite its subjective character, experienced as an objective feature of the world. There are, I believe, two ways to respond to this objection.

First, the objection equivocates on the meaning of 'disgust'. Insofar as SUBJECTIVISM construes disgust as being understood subjectively, it is committed to the claim that we do not experience our own disgust as an objective feature of the world. However, the projectivist response outlined above amounts to the claim that we sometimes *do* experience our disgust as an objective feature of the world. If there is such a mental state that is projected onto the world and then read off as an objective feature, this mental state cannot be the same as that disgust identified by SUBJECTIVISM. Disgust, as it is construed by SUBJECTIVISM, is not the kind of thing we experience as an objective feature of the world. Whatever kind of property moral wrongness turns out to be, the phenomenological thesis is committed to the claim that it must be the sort of thing that

we experience as an objective feature of the world. The property of moral wrongness and the property of being the object of my disgust cannot, therefore, be identical.

Second, the 'bent oar' objection described above misunderstands the argument that moral projectivism is a theory that has moral content. It is true that our experience of the oar in the example above does not entail that the oar is bent. Similarly, our experience of moral properties as objective does not entail that moral properties are objective. Our experience of the oar does, however, tell us that oars are the sorts of thing that appear bent when placed in water. If there were an oar that did not have this property, we could not continue to believe that oars appear bent when submerged in water and also believe that the object in question is in fact an oar. We would have to revise our theory about oars to in some way allow for the possibility of an oar that does not seem bent when submerged in water. Similarly, the phenomenological thesis is committed to the claim that SUBJECTIVISM is false, as the class of things SUBJECTIVISM identifies as being morally salient lack the property of being experienced as an objective feature of the world. In order to allow for the possibility of SUBJECTIVISM being true, the phenomenological thesis would have to admit the possibility that moral properties do not always seem to be objective features of the world, as some moral properties may be things that seem subjective. However, this admission would amount to a rejection of the phenomenological thesis. There is, then, no way to formulate the phenomenological thesis such that it lacks moral content.

This is yet another case that demonstrates how difficult it is to say anything philosophically illuminating about the nature of moral judgment without committing oneself to substantive moral propositions and thus disqualifying one's theories from

being empirically-confirmable. It may be true that most people, when pressed, would claim that moral properties are in many ways like objective features of the world (for example, many people would likely claim that moral truths are more objective than matters of taste). However, when philosophers make transitions from claims like this to claims about the nature of moral judgment *simpliciter* they often do not take notice of the fact that the newly articulated theory is one that has moral content. And, as I have argued, whenever a proposition has moral content it cannot be construed as an empirically-confirmable hypothesis.

VII - Psychological Realism

I have argued that at least three prominent theories about the nature of moral judgment cannot be confirmed or disconfirmed by empirical evidence because all of these theories presuppose commitments to substantive moral propositions. It is true that I have not discussed all possible empirically-informed theories of the nature of moral judgment, but the fact that three of the most prominent attempts all succumb to the same criticism suggests that my conclusion can be further generalized. The evidence I have provided strongly suggests that no philosophically interesting[40] theory of moral judgment must be grounded primarily in moral considerations. However, some philosophers have argued that empirical science must play some role in metaethical debates. John Doris is one

[40] I understand a 'philosophically interesting' theory of moral judgment to be one that competes with and contradicts the kinds of theories that are articulated and defended by philosophers. Some examples of such theories are neosentiementalism, projectivism, ideal observer theory, and relativism.

prominent proponent of a constraint on theories of moral judgment that he dubs 'psychological realism' (PR):

> PR: Ethical reflection should be predicated on a moral psychology bearing a recognizable resemblance to actual human psychologies. (Doris 2002 p. 112)

I will argue that PR cannot be appealed to in order to ground a morally neutral, empirical theory of moral judgment because when we try to re-formulate PR such that it establishes a connection between empirical evidence and metaethical theories, it too must be understood as having moral content.

Before I can turn to this argument, it is important to note that there are some cases where empirical evidence must inform metaethical theories. If a particular theory is committed to some empirically-verifiable claim which turns out to be false, that theory should be rejected. But this is obviously true only of theories that are explicitly committed to an empirical claim. If, for example, a theory of the nature of moral judgment is predicated upon the claim that all humans would give a specific response to a particular question, the discovery that not all humans give that response would serve to refute the theory. Another example might be a metaethical theory that is predicated on the hypothesis that all humans act exclusively for selfish motives. If a theory was committed to this claim, it could no doubt be refuted with empirical evidence. But very few philosophical theories of moral judgment explicitly include empirically-testable predictions like these. So why, then, have so many philosophers supported PR as a constraint on theories of moral judgment?

As I have shown above, many theories about the nature of moral judgment are phrased in such a way as to make them seem at first glance as though they should be empirically verifiable. But these theories are not best understood as empirical hypotheses, as no amount of empirical evidence can confirm or disconfirm such a theory without additional moral premises. The relevant difference, as I have been stressing, is between propositions of the two following forms:

1) Most people, when prompted in a survey, answer questions about properties like ‘wrongness’ in the same way that they answer similar questions about objective features of the world
2) Moral properties seem like objective features of the world

The former is empirically-confirmable but the latter, while it might look at first glance as if it would be confirmed by the empirical evidence in (1), actually has moral content insofar as it gives moral advice in some circumstances. And as I have argued, any proposition that has moral content cannot be formulated as an empirically-confirmable hypothesis. Yet, as I have shown above, philosophers sometimes move seamlessly between propositions like (1) and (2).

The mistake of equating empirical hypotheses with moral propositions is comparable to the mistake Moore dubbed the naturalistic fallacy. Both involve philosophers illicitly equating moral propositions with naturalistic ones. My arguments diverge from Moore’s primarily in how they demonstrate that this move is a mistake. Moore focuses on the tendency of philosophers to mistakenly identify normative properties with natural ones, whereas I argue that the fact that moral disagreement cannot

be explained as a disagreement over an empirical hypothesis implies that moral propositions cannot be construed as empirical hypotheses. As I have argued, the best available theory of the nature of moral content implies that propositions with the form of (2) must be understood as moral propositions.

But the argument I have put forward to this point only establishes that *some* attempts to apply empirical considerations to metaethical debates are mistaken, namely, those that equivocate between propositions like (1) and (2). It remains an open question, then, whether or not there will be any circumstances in which we require that metaethical theories answer to the empirical sciences of psychology and neuroscience. The attitude we ought to have towards PR therefore turns on this question.

I am not the first to address this question. Virginia Held (1998) has argued that if cognitive science develops a conception of moral judgment that conflicts with that developed and endorsed by ethicists, the scientific explanation should not be privileged:

> …those of us interested in ethics should insist, I think, on pursuing our agenda: finding a conception of mind compatible with what we understand and have a good reason to believe about moral experience. If this conception of mind is incompatible with cognitive science, so much the worse for cognitive science. (Held 1996 p. 73)

Held is here offering her opinion regarding what ethicists should do if we find that the theory of moral judgment that best accounts for our moral experience conflicts with a scientific conception of moral judgment. Her answer, as is obvious from the above text, is

that there is no reason to presume that the scientific view should have priority. This amounts to a rejection of PR on Held's part.

It is important to note that both Held and the proponents of PR are committed to the claim that empirically-supported theories of moral judgment *can* conflict with theories of moral judgment that are grounded in moral discourse. One consequence of the arguments from this chapter is that this is not usually the case. When philosophers take themselves to have formulated an empirically-confirmable theory of moral judgment, they have very often mistaken a moral proposition for an empirically-confirmable one or failed to notice the moral foundations of the theory in question. When this is the case and the putatively empirical proposition is in fact a moral proposition, no amount of empirical evidence could prove or disprove it absent additional moral premises. Since it is not possible for empirical evidence to contradict moral propositions without additional premises connecting them (as I argued above), conflict between theories of the nature of moral judgment and empirical evidence is therefore mostly, if not entirely, illusory.

One example of the illusion would be a philosopher claiming that accepting (1) but denying (2) (or vice versa) would be contradictory. It does seem, at first glance, as if the evidence in (1) should be relevant to determining whether (2) is true. However, someone could coherently and consistently hold to one and yet deny the other. This would be possible, for example, under the presupposition that most people do not have a correct understanding of moral experience. There is, under this presupposition, no conflict between a denial of (1) and an assertion of (2), nor is there a conflict between an assertion of (1) and denial of (2). Just as large groups of people can have faulty moral beliefs, they can also have a faulty understanding of the nature of the moral domain. If

this were true of the group discussed in (1), the answers they provide to questions about the phenomenology of moral judgment could be dismissed because those answers are predicated on a mistaken understanding of what class of judgments are moral judgments. It is therefore quite plausible to presume that some or many people could have false beliefs about the phenomenology of moral judgment. If these subjects have fundamentally misconstrued the nature of morality, then their answers to questions about their experience of the objects of these judgments cannot be trusted and would not serve as evidence in favour of or against (2). Despite an illusion to the contrary, (1) and (2) are not related to one another.

There are other examples of the illusion of a conflict between empirical evidence and moral theses. Some philosophers, such as Nichols, believe that empirical investigations into how people understand the nature of moral rules can resolve debates about what these rules are actually like. In order to be able to make this kind of argument, we would have to be able move from the morally neutral, empirical claim that:

(3) It has been empirically demonstrated that most people understand moral rules as being unconditionally obligatory, generalizable, and impersonal insofar as they stem from concepts of welfare, justice and rights.

to the moral claim that:

(4) Moral rules are unconditionally obligatory, generalizable, and impersonal insofar as they stem from concepts of welfare, justice and rights.

We cannot make this move without an intermediary premise. Such a premise would have to imply that the way people *do* conceive of the content of the moral domain is decisive of what the content of the moral domain actually is. A version of PR might serve as such a premise, but PR itself is too vague to do this work. In order to bridge the gap between (3) and (4), we require a more detailed, specific version of PR:

> PR*: If it has been empirically demonstrated that most people understand moral rules as having properties X, Y and Z, then moral rules have properties X, Y and Z.

PR* is not implied by PR, but it is the kind of premise that would be needed for empirical evidence like that found in (3) to impact philosophical inquiry into the nature of moral rules. Without such a premise, there is no connection between (3) and (4). I could, for example, accept (3) and yet deny (4) on the grounds that most people have a corrupt understanding of moral rules.

The fact that we need such a premise is demonstrative of exactly why theories of the nature of moral judgment that have moral content do not conflict with empirically-confirmable theories. Without additional premises, we have no way of arguing that we must include empirical evidence in debates about the nature of moral judgment.

Furthermore, this example demonstrates why philosophers cannot appeal to psychological realism or premises like PR* in order to ground a morally-neutral, empirical theory of the nature of moral judgment. This is because PR* must itself be understood as having moral content. PR* would give moral advice in the circumstance

wherein impartial considerations of justice, welfare and rights conflict with normative theories that prioritize caring for those with whom we are in concrete personal relationships or the production of virtue. If, in such a situation, I accept PR*, then PR* plays the role of a moral standard. PR* would tell me that the morally correct action in this case would be the one that conforms to the rule that has the properties most people attribute to moral rules. Under the presupposition that the morally-neutral proposition (3) is true, considerations of justice would win out. This example demonstrates that PR* identifies certain naturalistic features of the world as being uniquely salient to the justification of a theory of the nature of moral rules. Because of this, it must be understood as giving moral advice in certain circumstances and therefore has moral content. It is more abstract than most paradigmatic moral standards, but it is a moral standard nonetheless. PR* must therefore be understood as a principle with substantive moral content and thus cannot ground a morally neutral theory of the nature of moral judgment. Furthermore, I suspect that we could not invoke PR to connect propositions like (3) and (4), or (1) and (2), without filling PR out in more detail, as a version of PR*. I will not defend this claim at this time, but will treat the example I gave above regarding PR* to be satisfactory to demonstrate its plausibility. PR does not itself tell us when a particular piece of empirical evidence is relevant to the assessment of a particular metaethical theory, all it tells us is that at least sometimes, it must be relevant. We cannot apply PR without supplementing it with additions like PR*. And additions of this sort, insofar as they serve to connect propositions (3) and (4), will have moral content.

My answer to questions about the status of PR therefore cannot be the same as Held's, as I reject one presupposition that she has in common with the proponents of

empirically-informed metaethics. Whereas she thinks that it is possible for empirical and moral accounts of moral judgment to conflict, I have argued that there are no such conflicts. The theses that are thought to be empirical usually wind up being disguised versions of moral propositions. Just as no body of empirical evidence could prove or disprove any substantive moral proposition, there is no such body of evidence that could independently prove or disprove a theory of the nature of moral judgment. In order to incorporate empirical evidence into metaethical debates about the nature of moral judgment, one would need to rely on further premises that must also have substantive moral content. In other words, our choice of whether or not to include empirical evidence into metaethical debates must be grounded on further moral assumptions.

But, despite these considerations, one could argue that there is still some role for empirical considerations in some investigations regarding moral judgment. Let us say, for example, that we have accepted the definition of morality that Turiel borrowed from Aristotle, Dworkin and Gewirth. Recall that Turiel defines social conventions as “part of constitutive systems and our shared behaviours (uniformities, rules) whose meanings are defined by the constituted system in which they are embedded” (Turiel, Killen, Helwig 1987 p. 169) and moral rules as “unconditionally obligatory, generalizable, and impersonal insofar as they stem from concepts of welfare, justice and rights” (ibid, pp. 169-170). Once we have settled on this definition, it would be possible to engage in empirical research of, for example, when children begin to make moral judgments or the best way to approach the moral education of children. However, such an investigation would not be a part of moral philosophy, but rather developmental psychology. I agree with Held that “we ought to have inquires we could call moral sociology, moral

psychology, moral economics, moral political sciences, and so on" (Held 1996 p. 69), but it would be a mistake to think of these research programs as engaging in the same kind of project as moral philosophy. While this type of research could answer some potentially interesting questions, it would not answer the questions debated by moral philosophers about the nature of moral judgment. It is only if we previously accept the theories of moral judgment offered by some philosophers that we can apply scientific findings to questions about morality. This type of experimental work does not, then, constitute an empirical investigation into the nature of moral judgment, but rather an application of the conceptual results from a philosophical, morally-loaded theory of moral judgment to an empirical project.

None of this is to say, however, that empirical scientists should not continue to do what they do. Rather, I have been arguing that philosophical theories of moral judgment *simpliciter* cannot be construed as empirically-confirmable hypotheses. Many psychological and neuroscientific studies do not formulate hypotheses about the nature of moral judgment *simpliciter*, but rather seek to answer other interesting questions. Jonathan Haidt, for example, has performed studies aimed at explaining the methods of reasoning that are used in justifying specific kinds of actions. In these studies, Haidt offers subjects particular scenarios and then asks whether or not the act in question should be done. One such case is the question of whether or not an individual should cook and eat a chicken that had just been used as a masturbatory aide (Haidt et al. 1993). Based on the responses subjects give to questions like these, Haidt concludes that "affective reactions to the stories…were better predictors of their moral judgments than were their claims about harmful consequences" (Haidt 2001 p. 817).

A similar methodology is also used by Joshua Greene. In one of Greene's studies, he presents subjects with a few different scenarios. Some involve up close and personal cases of harming a person in order to save others, while others involve less personal cases in which the subject has no physical contact with the individual that is sacrificed (the commonly discussed footbridge and trolley cases). Other cases provided to subjects do not involve this kind of act, but rather concern putatively 'non-moral' instances of deciding, for example, which of two coupons to use (Greene 2001 pp. 2105-2106). Greene then concludes that "from a psychological point of view, the crucial difference between the trolley dilemma and the footbridge dilemma lies in the latter's tendency to engage people's emotions in a way that the former does not" (ibid).

What both Greene and Haidt's work have in common is a shared methodological approach. Their method consists in presenting subjects with token judgments and then recording the types of justifications they give and observing the neural processes that were activated while thinking about them. Because they design their studies this way, these psychologists do not initially presume that there is a distinct category of moral judgment, nor does their experimental methodology require that they explicitly define at the outset the difference between moral and non-moral judgments. While they wind up drawing general conclusions about the psychological and neurological character of moral judgments, they do not commit themselves to any robust definition of them:

> Morality is probably not a 'natural kind' in the brain. Just as the ordinary concept of memory refers to a variety of disparate cognitive processes (working memory, episodic memory, motor memory, etc.), we believe that the ordinary concept of moral judgment refers to a variety of more fine-

> grained and disparate processes, both ‘affective’ and ‘cognitive’. (Greene & Haidt 2002 p. 523)

Insofar as this methodology does not presuppose any morally-loaded theses or aim to confirm, by empirical means, any theory of the content and nature of moral judgment *simpliciter*, the arguments I have raised above do not apply to these theories.[41]

However, any attempts to apply these empirical findings to philosophical debates about the nature and content of moral judgment will succumb to the criticisms I raised above. I do not take myself to be criticizing scientists, but rather philosophers who attempt to appeal to empirical research in order to defend their preferred account of moral judgments without acknowledging that in doing so, they commit themselves to substantive moral commitments. Empirical research has its place, but, to paraphrase Shafer-Landau, the nature of moral judgment will not be determined by those folks who wear lab coats, but rather by those who think philosophically.[42]

---

[41] This claim only applies to the works by Greene and Haidt that are here cited. In other places, these psychologists have predicated their investigations on substantive assumptions about the content of morality. See for example Haidt 2007.

[42] For the sake of clarity, it is important for me to explain how I understand this claim. I acknowledge that empirical considerations *may* play some role in debates about the nature of moral judgment. Just as we must think about many empirical considerations when engaging in moral deliberation (e.g., we must know that we are accurately characterizing the situations under discussion before we can be certain that our moral judgments about those situations are correct), philosophers may discover that there are good moral reasons to pay attention to empirical findings in moral psychology. Nonetheless, just as the fact that empirical considerations figure in moral deliberation does not undermine the autonomy of morality, I believe that we can still legitimately claim that empirical findings are not relevant to (or cannot significantly inform) debates about the nature of moral judgment. This is because empirical considerations do not seem to independently confirm or disconfirm any theory about the nature of moral judgment. Empirical considerations may wind up playing some role in our thought about these questions, but theories about the nature of moral judgment like those discussed above can be rejected solely on the basis of their moral commitments. It is in this sense that I say that empirical considerations are not relevant to, or do not inform, these debates. Whenever we find that empirical considerations are relevant to moral reasoning, it is only because additional moral premises identify those considerations as being relevant. Their relevance is therefore always secondary, or derivative. This is what I mean when I say that empirical evidence cannot determine the nature of moral judgment. Thank you to William Harper and Charles Jones for pointing out that there is a need for clarity on this point.

VIII - Moral Methods

Finally, it is worth asking how we should go about developing and defending theories about the nature of moral judgment. If we cannot appeal to empirical research, what methodology should we employ? My answer is that we must approach these theories as we would any other moral theory. We must engage in moral discourse with our peers, think through our considered moral judgments and attempt to achieve a reflective equilibrium between our particular moral judgments and our more abstract theories. We must, in short, treat theories about the nature of moral judgment as we would any other thesis that must be construed as being internal to moral discourse. Philosophers who have understood these theories as being held hostage to empirical findings have indentified the wrong body of considerations as being relevant to their confirmation or disconfirmation.

For example, there are good moral reasons to think that moral rules are objective in a robust sense (See Parfit 2011, pp. 74-80 or Enoch 2010 for an argument to this effect). There is also good evidence for the related claim that moral properties are the sorts of things that seem to be objective features of the world. As I have argued, this thesis must be understood as a moral claim, namely, the moral claim that only certain kinds of properties (those that seem to be objective features of the world) are uniquely morally salient. Anyone who does not understand moral properties as the sorts of things that seem objective has misconstrued the nature of moral discourse and thus has a faulty conception of their own moral phenomenology. Furthermore, we can raise moral

arguments against relativists such as Prinz and Harman. Even if certain cultures have sentimental dispositions that reflect positively on morally atrocious actions such as the infibulation of young females, I see no contradiction in claiming that they morally ought to do otherwise and that they ought to have different attitudes towards this terrible practice. This amounts to a rejection of relativism for moral reasons. There are also moral reasons to endorse some features of Turiel's characterization of morality. His assertion that morality is unconditional, for example, is comparable to Kant's pronouncement that "a law, if it is to hold morally, that is, as a ground of an obligation, must carry with it absolute necessity" (Kant 1785 p. 44). These claims, if they are to be proven, must be proven on the basis of their moral content. Some of the theories canvassed in this and previous chapters are confirmed by moral evidence while others are disconfirmed by such evidence. The confirmation or disconfirmation of these theories is completely independent from the findings of empirical sciences.

However, we cannot confirm any of the theories discussed above unless we have accepted an account of what serves as evidence for or against a moral proposition. I believe that a number of theories about this are consistent with everything I have said above. For example, nothing I have said would rule out a foundationalist moral epistemology according to which certain self-evident moral propositions justify all moral propositions derived from them (although I admit that I do not know which propositions are the best candidates for being considered self-evident). It is also consistent with Allan Gibbard's theory according to which moral deliberation is understood by analogy to contingency planning (See, for example, Gibbard 2003 pp. 53-58). The most plausible option, however, is some version of reflective equilibrium. Narrow reflective

equilibrium, according to which we engage in moral discourse by seeking an equilibrium between only our particular moral beliefs and moral standards, could be appealed to in order to ground any of the theories about the nature of moral judgment that I have addressed above. We could also appeal to a wider version of reflective equilibrium, according to which we must establish an equilibrium between a much wider set of beliefs (about, for instance, the role of morality in society, the nature of a person, and so on) as long as this wide reflective equilibrium does not take itself to be grounding morality in non-moral considerations.[43] This should not be a problem, as many of the considerations that are thought to be important to include within a wider reflective equilibrium would, according to the wider conception of moral content I argued for in chapter one, be best interpreted as being internal to moral discourse.

While my view is consistent with any of these methods of theory selection in ethics, I believe that the best view is a compromise between the narrow and wide versions of reflective equilibrium that explicitly recognizes this last point. A conception of reflective equilibrium that asks us only to balance our first-order moral judgments with a set of normative moral standards and principles is oversimplified and risks collapsing into a form of intuitionism (Daniels 1979). Many other considerations and arguments can and should impact which normative theories we endorse. For example, consider arguments about the nature of personhood or the concept of the person. An argument that is purported to establish the immorality of abortion, for example, might rely on the premise that a human life begins at conception and that all human lives are persons. Such

[43] Proponents of wide reflective equilibrium think this can be accomplished. Norman Daniels, for example, does not think that this process "constitutes a reduction of the moral…to the nonmoral." (Daniels 1979 p. 260).

arguments were historically quite prevalent in the early years of the abortion debate (English 1975). One way of refuting this argument would be to rehearse some of the rigorous philosophical arguments about personhood in order to show that the concept of being a person is not co-extensive with the concept of being a human being (See, for example, Frankfurt 1971 and Dennett 1976). Such an argument would therefore be seen to be relevant to morality in just the way Daniels describes:

> We do *not* simply settle for the best fit of principles with judgments, however, which would give us only a *narrow* equilibrium. Instead, we advance philosophical arguments intended to bring out the relative strengths and weaknesses of the alternative sets of principles (or competing moral conceptions). These arguments can be construed as inferences from some set of relevant background theories (I use the term loosely.) (Daniels 1979 p. 258)

However, incorporating these broader philosophical arguments into our reflective equilibrium does not amount to an admission that morality is reducible to non-moral considerations. There are, instead two possibilities of how to understand this incorporation that maintains the autonomy of morality.

First, it is possible that the concept of personhood under discussion in this body of literature is itself a moral concept. Daniel Dennett, for example offers a similar answer when he addresses the question of why it is so hard to articulate sufficient conditions for personhood: "simply because the concept of a person is, I have tried to show, inescapably normative" (Dennett 1976 p. 193). It might be the case that the concept of personhood is itself a normative concept in a distinctively moral way. If this is the case, then

philosophical arguments about the nature of personhood are some of the arguments that we may have to take into account when trying to assess moral propositions.

Second, even if it is not the case that personhood is best understood as a moral concept, our decision to include philosophical arguments about the nature of a person in our reflective equilibrium must be justified by an additional moral premise. For example, before we could conclude that the nature of personhood is relevant to our choice of normative theories, we would need to endorse a proposition like the following: ‘the nature of personhood is relevant to morality.’ This premise is what gives us reason to incorporate considerations regarding personhood into our reflective equilibrium. Just as the moral judgment that only pleasure is morally salient gives us reason to incorporate considerations about the nature of pleasure into our reflective equilibrium, the judgment that the nature of a person is morally relevant similarly requires us to consider philosophical arguments about personhood when discussing normative moral theories. The incorporation of such background theories into our reflective equilibrium does not reduce morality to non-moral considerations because we still require moral judgments in order to tell us which philosophical arguments to include and which to exclude.[44] The truth of ethical hedonism does not imply that the proposition ‘the consumption of ice cream induces pleasurable states’ has moral content, but hedonism does require us to think about what sorts of things induce pleasurable states in order to arrive at true moral propositions.

---

[44] While I believe that an insufficient number of philosophers have embraced this point, I do not claim that it is entirely original. It has, I believe been held by other influential philosophers. For example, my thesis seems to be consistent with John Rawls’s claim that “moral conceptions regard persons differently and prize different aspects of their nature. So although every conception employs a criterion of identity that recognizes the results of the philosophy of mind, each may specialize its criterion to fit the requirements of a particular moral order and conception of the person. The comparative study of these matters belongs to moral theory and takes us beyond the philosophy of mind” (Rawls 1975 p. 17).

This conception of reflective equilibrium is wide in the sense described by Daniels, but is also narrow insofar as it recognizes that the process of trying to arrive at a reflective equilibrium is shaped and directed by moral considerations. I prefer this version of wide reflective equilibrium because it treats the question of whether or not the findings of empirical disciplines are ever relevant to moral considerations as a moral question. I argued above that this is the correct approach to take towards psychological realism. If we endorse psychological realism, then the empirical data plays the same role in our moral theorizing as do theories of personhood and pleasure in the cases discussed above. This role may prove to be extremely important and central to much of moral thought, but it is nonetheless contingent upon individual moral claims about when empirical moral psychology is relevant. The question of whether or not psychological realism is true must therefore be understood as a moral question. This is true for many other theses that are often taken by philosophers to warrant the inclusion of empirical psychology into ethical arguments.

Consider, for example, the dictum that ought implies can. This dictum is usually taken to mean that it is impossible for one to be morally obliged to do something which one is incapable of doing. If this thesis is correct, then we would have to incorporate empirical considerations about what people are or are not capable of doing into our moral reflection. But it is important to note that the question of whether or not 'X morally ought to Y' always implies 'X can Y' must itself be understood as a moral question. The idea that we cannot be obliged to do the impossible is a distinctively moral idea, one that could be rejected for moral reasons. For example, a normative moral theory that required all humans to remain completely calm and not experience any affective emotions would

give moral advice that is inconsistent with the dictum ‘ought implies can’ (under the very plausible presupposition that it is not possible for all humans to do so). This is sufficient for concluding that the dictum ‘ought implies can’ has moral content. This is one example of a case wherein extramoral concerns prove to be relevant to our moral thought, but it is clear that their relevance is entirely contingent on the moral judgment that they are so.

The version of wide reflective equilibrium that I have defended is preferable to any of the alternative theories that try to explain theory choice in ethics. First, it has the benefit that Daniels attributed to wide theories of reflective equilibrium of circumventing some of the problems for foundational intuitionism. Such theories do not rely on any set of judgments serving as an epistemic foundation insofar as all beliefs within the equilibrium are susceptible to revision and can potentially be abandoned (Daniels 1979 p. 266). Second, it conforms well with what I have said about the methodology of metaethics. Many things may prove to be relevant to moral deliberation, but this relevance is always contingent upon further moral judgments. In a very important way, only moral propositions can be appealed to in order to serve as evidence against other moral propositions. This position, which is accurately thought of as the thesis that ethics is autonomous, is best explained and supported by the kind of reflective equilibrium I have here described.

However, my thesis that theories of moral judgment cannot be construed as empirical hypotheses is not contingent upon the success of this last argument in favour of wide reflective equilibrium. It is my view that this is the most appropriate method for rationally selecting normative moral theories, but I am open to the possibility that other

methodologies are equally viable. The most important conclusion of this chapter is that once we have settled on such a methodology, we use it for all moral propositions, including theories about the nature of moral judgment.

Conclusion

I have argued that theories about the nature of moral judgment must be understood as being internal to moral discourse. This means, amongst other things, that these theories are confirmed or disconfirmed largely by moral considerations, rather than empirical considerations. We must therefore evaluate them in the same way that we evaluate normative moral theories like utilitarianism, virtue ethics and deontology. There are a number of theories about how we should go about doing this. I have argued that a version of wide reflective equilibrium is the most plausible of these, but my central thesis that theories of the nature of moral judgment must be understood as moral propositions does not depend on the success of this argument.

My argument for the controversial conclusion that theories of moral judgment have moral content began with the development of a criterion for determining whether or not a proposition has moral content, which I dubbed MORAL. I showed that my proposal avoided many of the criticisms that could be raised against alternative theories. I then demonstrated that many influential metaethicists have failed to notice the fact that their preferred theories about the nature of moral judgment actually have moral content and I responded to some possible objections to this claim. I began the final chapter by articulating an argument in favour of the claim that moral propositions cannot be identical to or reducible to empirically-confirmable hypotheses. This argument is largely inspired by recent refinements of G.E. Moore's arguments against naturalism, but it is also an original contribution to this body of literature insofar as it substantially reframes

the argument and expands on other recent versions. Finally, I argued that theories of the nature of moral judgment cannot be confirmed or disconfirmed by empirical evidence, but rather must be confirmed or disconfirmed predominantly by moral considerations.

I have advocated a fairly radical re-thinking of certain features of metaethical practice. Despite the fact that many philosophers endorse empirically-minded metaethics, I believe these arguments demonstrate that empirical theories of moral judgment cannot serve as an archimedean point from which we can construct a philosophical theory of moral judgment.

Some philosophers may find this conclusion dismaying. The empirical turn in philosophical moral psychology is not meant to undermine metaethics, but rather to provide it with a more secure foundation. Presumably if a metaethical theory can be confirmed by empirical evidence, this would count in its favour. But should we despair over the fact that this particular class of theories cannot be empirically confirmed?

I can see why some philosophers might be inclined to think so. Various forms of naturalism are extremely attractive because of their parsimony and ontological modesty. It is therefore understandable that philosophers want to extend this naturalism to a variety of philosophical questions. But nothing I have said is inconsistent with the naturalist's goal of being parsimonious and limiting the number of mysterious or unexplained properties in our ontology. My favoured form of non-naturalism does not posit non-natural properties, but rather argues that our moral concepts cannot be understood in naturalistic terms. This does not presume the existence of ontologically strange values, nor does it preclude the possibility that even the most robust forms of naturalism are true.

All we need in order to argue in favour of a theory of the nature of moral judgment is a workable account of how we can go about rationally deliberating about the faults and merits of particular moral propositions. This is not a particularly lofty or idealistic goal. Moral deliberation and reasoning play a large role in most people's lives. My argument here is just that this type of reasoning extends somewhat further than many philosophers have recognized or acknowledged. If it is true that moral questions are meaningless or that all moral disagreement is founded on a mistake, then these problems would also infect our attempts to defend an account of the nature of moral judgment. However, even if it could be established that the most nihilistic metaethical theory is in fact accurate and all moral claims are incoherent, or all moral intuitions are suspect, this would not undermine my argument. All it would mean is that theories about the nature of moral judgment fall within the class of the meaningless judgments. We would have to pass over debates about these theories in silence just as we would debates in normative ethics.

However, I am not convinced that our position is so bleak. In order to be justified in making moral judgments and endorsing moral beliefs, we do not need to defend a form of full-blooded moral realism that is committed to the mysterious Platonic existence of moral facts and values. All we need to do is be able to offer an account of how it is that we can rationally make use of moral discourse in order to regulate our conduct, discuss our attitudes and deliberate about how we should behave. There are, nonetheless, a number of theories that completely denounce participation in moral discourse. A complete taxonomy of all the possible views that aim to delegitimize moral thinking would take me too far afield, as would a full defense of morality from all these possible

detractors. However, it is important to note that any such position would have to be predicated on an account of the nature of moral judgments. Without this starting point, it would not be at all clear what this position is asking us to denounce. And if these starting assumptions must be understood as having moral content, then the theory must be committed to the kinds of beliefs that it aims to dismiss. Most of the theories that are skeptical of morality or moral intuitions *simpliciter* that I have encountered succumb to a criticism of this type, although I admit that for now I leave this claim undefended.

It is of course possible that a form of skepticism could escape this line of attack. But nonetheless, nothing I have said in the above chapters would be refuted by such a skeptical thesis. The skeptic could conclude that there is no right answer to the question ‘what is a moral judgment?’ But this question would remain, as I have maintained, a moral one.

Bibliography

Alm, David. “Moral Conditionals, Noncognitivism and Meaning.” *The Southern Journal of Philosophy* 38 (2000): 355-377.

Ayer, A.J. *Language, Truth, and Logic.* New York: Dover, 1936/1952.

Blackburn, Simon. “Attitudes and Contents.” *Ethics* 98 (1988): 501-517.

Blackburn, Simon. “Truth, Beauty and Goodness.” *Practical Tortoise Raising and Other Philosophical Essays*. Oxford: Oxford University Press, 2010: 26-46.

Blair, R.J.R. “A Cognitive Developmental Approach to Morality: Investigating the Psychopath.” *Cognition* 57 (1995): 1-29.

Blair, R.J.R, Jones L. Clark, F. and Smith, M. “The Psychopathic Individual: A Lack of Responsiveness to Distress Cues?” *Psychophysiology* 34 (1997): 192-8.

Blair, James, Mitchell, Derek, Blair, Karina. *The Psychopath: Emotion and the Brain*. Oxford Blackwell Publishing, 2005.

Brandt, Richard. “The Emotive Theory of Ethics.” *The Philosophical Review* 59 (1950): 305-318.

Brink, David. *Moral Realism and the Foundations of Ethics*. New York: Cambridge University Press, 1989.

Dancy, Jonathan. “Two Conceptions of Moral Realism.” *Proceedings of the Aristotelian Society* 60 (1986): 167-187.

Daniels, Norman. “Wide Reflective Equilibrium and Theory Acceptance in Ethics.” *The Journal of Philosophy* 76 (1979): 256-282.

Darwall, S. Gibbard, A. and Railton, P. “Toward fin de siècle ethics: Some Trends.” *The Philosophical Review* 101 (1992): 115-189.

Dennett, Daniel. “Conditions of Personhood.” *The Identities of Persons*. Ed. A.O. Rorty. Berkeley: University of California Press, 1976: 175-196.

Doris, John M. *Lack of Character*: *Personality and Moral Behaviour*. Cambridge: Cambridge University Press, 2002.

Dreier, James. “Meta-ethics and Normative Commitment.” *Philosophical Issues* 12

(2002): 241-263.

Dworkin, Ronald. "Truth and Objectivity: You'd Better Believe it." *Philosophy and Public Affairs* 25 (1996): 87-139.

Ehrenberg, Kenneth. "Archimedean Metaethics Defended." *Metaphilosophy* 39 (2008): 508-529.

English, Jane. "Abortion and the Concept of a Person." *The Canadian Journal of Philosophy* 5 (1975): 233-243.

Enoch, David. "How Objectivity Matters." *Oxford Studies in Metaethics 5.* Ed. R. Shafer-Landau. Oxford: Oxford University Press, 2010: 111-152.

Fantl, Jeremy. "Is Metaethics Morally Neutral?" *Pacific Philosophical Quarterly* 87 (2006): 24-44.

Foot, Philippa. *Virtues and Vices and Other Essays in Moral Philosophy.* Berkeley and Los Angeles: University of California Press. 1978.

Frankena, "Moral Philosophy Mid-Century." *Philosophical Review* 60 (1951): 44-55.

Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68, 1971: 5-20.

Geach, P.T. "Assertion." *The Philosophical Review* 74 (1965): 449-465.

Gewirth, A. "Meta-ethics and Normative Ethics." *Mind* 69 (1960): 187-205.

Gibbard, Allan. *Wise Choices, Apt Feelings.* Cambridge: Harvard University Press, 1990.

Gibbard, Allan. *Thinking How to Live*. Cambridge: Harvard University Press, 2003.

Gibbard, Allan. "Normative Properties." *Metaethics after Moore:* Ed. T. Horgan & M. Timmons. Oxford: Oxford University Press, 2006a: 319-338.

Gibbard, Allan. "Moral Feelings and Moral Concepts." *Oxford Studies in Metaethics Vol. 1.* Ed. R. Shafer-Landau. Oxford: Oxford University Press, 2006b: 195-216.

Gilligan, Carol. *In a Different Voice: Psychological Theory and Women's Development.* Cambridge: Harvard University Press, 1983.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., & Cohen, J.D. "An

fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (2001): 2105-2108.

Greene, J. and Haidt, J. "How (and where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6 (2002): 517-523.

Haidt, J. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108 (2001): 814-834.

Haidt, J. "The New Synthesis in Moral Psychology." *Science* 316 (2007): 998-1002.

Harman, Gilbert. "Moral Relativism Defended." *The Philosophical Review* 84 (1975): 3-22.

Held, Virginia. "Whose Agenda? Ethics versus Cognitive Science." *Mind and Morals: Essays on Ethics and Cognitive Science*. Eds: L. May, M. Friedman and A. Clark. Cambridge: MIT Press,1996: 69-88.

Horwich, Paul. *Truth*. Cambridge MA: Basil Blackwell Ltd., 1990.

Hume, David. *A Treatise of Human Nature*. Oxford: Clarendon Press, 1739/1974.

Hursthouse, Rosalind. "Virtue Theory and Abortion." *Philosophy and Public Affairs* 20 (1991): 223-246.

Johnson, Oliver A. "Rightness, Moral Obligation, and Goodness." *Journal of Philosophy* 50 (1953): 597–608.

Joyce, Richard. *The Evolution of Morality*. Cambridge: MIT Press, 2006.

Joyce, Richard. "Is Moral Projectivism Empirically Tractable?" *Ethical Theory and Moral Practice* 12 (2009): 53-75.

Kant, Immanuel. "Groundwork for the Metaphysics of Morals." *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Trans & Ed. Mary J. Gregor. Cambridge: Cambridge University Press, 1785/1996: 37-108.

Karmo, Toomas. "Some Valid (but no Sound) Arguments Trivially Span the 'Is'-'Ought' Gap." *Mind* 97 (1988): 252-257.

Kelly, D. Stitch, S. Haley, K. J. Eng, S. J. Fessler, D.M.T. "Harm, Affect, and the Moral/Conventional Distinction." *Mind and Language* 22 (2007):117-131.

Korsgaard, Christine. *The Sources of Normativity.* Cambridge: Cambridge University Press, 1996.

Kirchin, Simon. "Ethical Phenomenology and Metaethics." *Ethical Theory and Moral Practice* 6 (2003): 241–264.

Kramer, Matthew. *Moral Realism as a Moral Doctrine*. West Sussex: Wiley-Blackwell, 2009.

Loeb, Don. "The Argument From Moral Experience." *Ethical Theory and Moral Practice* 10 (2007): 469–484.

Mackie, J.L. *Ethics: Inventing Right and Wrong*. London: Penguin, 1977.

McDowell, John. "Values and Secondary Qualities." *Morality and Objectivity*. Ed. T. Honderich. London: Routledge & Kegan Paul, 1985: 166-180.

McNaughton, David. *Moral Vision: An Introduction to Ethics.* New York: Basil Blackwell Inc, 1988.

McPherson, Tristram. "Metaethics and the Autonomy of Morality." *Philosopher's Imprint* 8 (2008): 1-16.

Miller, Alexander. *An Introduction to Contemporary Metaethics*. Malden: Polity, 2003.

Moore, G.E. *Principia Ethica*. Cambridge: Cambridge University Press, 1903.

Mothersill, Mary. "Moral Philosophy and Meta-Ethics." *The Journal of Philosophy* 49 (1952): 587-594.

Nichols, Shaun. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press, 2004.

Nichols, Shaun. "Sentimentalism Naturalized." *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. Cambridge: MIT Press, 2008: 255-274.

Nietzsche, Friedrich. *On the Genealogy of Morals*. Trans. Walter Kauffman and R.J. Hollingdale. New York: Vintage Books. 1887/1967.

Olafson, F.A. "Meta-ethics and the Moral Life." *Philosophical Review*: 65 (1956): 1956.

Parfit, Derek. *On What Matters: Volume 1.* Oxford: Oxford University Press, 2011.

Pickard-Cambridge, W. A. "Two Problems About Duty (I.)" *Mind* 41 (1932): 72–

96.

Pigden, Charles. "Logic and the Autonomy of Ethics." *Australasian Journal of Philosophy* 67 (1989): 127-151.

Pigliucci, M. and Kaplan, J. *Making Sense of Evolution: The Conceptual Foundations of Evolutionary Biology*. Chicago: University of Chicago Press, 2006.

Prinz, Jesse. *The Emotional Construction of Morals*. Oxford: Oxford University Press, 2007a.

Prinz, Jesse. "Can Moral Obligations Be Empirically Discovered?" *Midwest Studies in Philosophy* 31 (2007b): 271-291.

Prinz, Jesse. "Is Morality Innate?" *Moral Psychology Vol. 1: The Evolution of Morality: Adaptations and Innateness*. Ed. W. Sinnott-Armstrong. Chicago: University of Chicago Press, 2008: 367-406.

Prior, A.N. "The Autonomy of Ethics." *Australasian Journal of Philosophy* 38 (1960): 199-206.

Railton, Peter. "Moral Realism." *The Philosophical Review* 95 (1986): 163-207.

Rawls, John. "The Independence of Moral Theory." *Proceedings and Addresses of the American Philosophical Association* 48 (1975): 5-22.

Ross, W.D. *The Right and the Good*. Oxford: Oxford University Press, 1930.

Ross, W.D. *Foundations of Ethics*. Oxford: Oxford University Press, 1939

Scanlon, T.M. *What We Owe to Each Other*. Cambridge: Harvard University Press, 1998.

Shafer-Landau, Russ. *Moral Realism: A Defence.* New York: Oxford University Press, 2003.

Shafer-Landau, Russ. "Ethics as Philosophy: A Defense of Ethical Non-Naturalism." *Metaethics after Moore:* Ed. T. Horgan & M. Timmons. Oxford: Oxford University Press, 2006: 209-232.

Smith, Michael. *The Moral Problem.* Oxford: Blackwell Press, 1994.

Stalnaker, Robert. "Assertion." *Syntax and Semantics*: Ed. P. Cole. New York: Academic Press, 1979: 315-332.

Stevenson, C.L. "The Emotive Meaning of Ethical Terms." *Mind* 46 (1937): 14-31.

Sumner, L.W. “Normative Ethics and Metaethics.” *Ethics* 77 (1967): 95-106.

Taylor, P.W. “The Normative Function of Meta-Ethics.” *Philosophical Review* 67 (1958): 16-32.

Tenenbaum, Sergio. “Ethical Internalism and Glaucon’s Question.” *Nous* 34 (2000): 108-130.

Tiberius, Valerie. “The Practical Irrelevance of Relativism.” *Analysis* 72 (2009): 722-731.

Toulmin, Stephen. *The Place of Reason in Ethics*. Chiacgo: University of Chicago Press, 1950.

Turiel, Elliot. *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press, 1983.

Turiel, E., Killen, M., Helwig C. “Morality: Its Structure, Functions and Vagaries.” *The Emergence of Morality in Young Children*. Eds. Kagan & Lamb. Chicago: University of Chicago Press, 1987: 155-244.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge: Harvard University Press, 1985.

Williams, Bernard. “Internal and External Reasons.” *Moral Luck*. Cambridge: Cambridge University Press (1981): 101-113.

Curriculum Vitae

Name:	Patrick Clipsham
Post-secondary Education and Degrees:	Wilfrid Laurier University Waterloo, Ontario, Canada 2003-2007 B.A. Honours Philosophy with Distinction
	The University of Waterloo London, Ontario, Canada 2007-2008 M.A. Philosophy
	The University of Western Ontario London, Ontario, Canada 2008-2012 Ph.D. Philosophy
Honours and Awards:	2008-2009 - Ontario Graduate Scholarship (declined)
	2008-2011 – Canada Graduate Scholarship
	2010 – Michael Smith Foreign Study Supplement
	2011 – Ontario Graduate Scholarship
Related Work Experience:	Teaching Assistant The University of Western Ontario 2008-2011
	Research Assistant The University of Western Ontario 2010-2012

Publications:
'Reasons and Refusals: The Relevance of Moral Distress.' *International Journal of Applied Philosophy* 26 (2012): 105-118.

Entry on 'Game Theory' *Encyclopedia of Utilitarianism,* eds. James Crimmins and Douglas Long (Bristol: Thoemmes/Continuum, forthcoming).