6-8-2007

Naturalism and Moral Realism

James Sias

Follow this and additional works at: http://scholarworks.gsu.edu/philosophy_theses

Recommended Citation
http://scholarworks.gsu.edu/philosophy_theses/21

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
ABSTRACT

My aim is to challenge recent attempts at reconciling moral realism and naturalism by pushing ethical naturalists into a dilemma. According to one horn of the dilemma, ethical naturalists must either (a) build unique facts and properties about divergent social structures (or varying moral belief systems) into their subvenient sets of natural facts and properties, and so jeopardize the objectivity of moral truths, or (b) insist, in the face of all possible worlds in which people have different moral beliefs than ours, that they are all mistaken—this despite the fact that the belief-forming mechanism responsible for their moral beliefs was never concerned with the truth of those beliefs in the first place. This will bring me to suggest that moral properties might only weakly supervene upon natural phenomena. But, according to the other horn of the dilemma, weak supervenience is a defeater for moral knowledge.

INDEX WORDS: Naturalism, Moral realism, Ethical naturalism, Moral properties, Supervenience, Darwinism, Justification, Reliabilism
NATURALISM AND MORAL REALISM

by

JAMES M. SIAS III

A Thesis submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2007
NATURALISM AND MORAL REALISM

by

JAMES M. SIAS III

Major Professor: Andrew Altman
Committee: George Rainbolt
Eddy Nahmias

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2007
To my wife,
Molly Campbell Sias,
and our son,
James M. Sias IV.
ACKNOWLEDGEMENTS

This thesis is the expression of ideas I’ve had for a few years now, but there is no way that it would have ever become what it is now without the help of my thesis director, Dr. Andrew Altman, and my thesis committee members, Dr. George Rainbolt and Dr. Eddy Nahmias. I would especially like to thank Dr. Altman, not only for his insights into the subject matter of this thesis, but also for what he’s taught me about the very project of doing philosophy. I’m also very grateful for Dr. Mark Linville, who not only played a foundational role in my decision to pursue philosophy, but also had many helpful things to say about this thesis during its construction. And finally, I want to thank Sean Martin and Justin Coates for ideas shared over beers, and my wife Molly for her encouragement.
TABLE OF CONTENTS

Acknowledgements v

Chapter One: An Introduction 1
  1.1 What “moral realism” is 6
  1.2 What “naturalism” is 8
  1.3 What “ethical naturalism” is 11

Chapter Two: Darwinism and Strong Moral Supervenience 12
  2.1 Weak and strong supervenience 13
  2.2 Dependence and relevance 15
  2.3 Darwinism and ethical naturalism 16
  2.4 Evolution and the formation of our moral beliefs 17
  2.5 A Darwinian dilemma for strong moral supervenience: the first horn 20
      2.5.1 Defending the first horn 23
  2.6 The second horn of the dilemma 24
      2.6.1 Defending the second horn 25
  2.7 Conclusion 26

Chapter Three: Weak Moral Supervenience and Justification 28
  3.1 Weak supervenience and the justification of moral beliefs 30
  3.2 Alternative hypotheses arguments 33
  3.3 Moral reliabilism 37
  3.4 Conclusion 40

References 41
CHAPTER ONE: AN INTRODUCTION

Referring to the realism/anti-realism debate in metaethics, Geoffrey Sayre-McCord writes, “The problem has always been to make sense of these [moral] convictions in a way that does justice to morality’s apparent importance without engaging in metaphysical flights of fancy.” It is a debate that came into sharp focus at the beginning of the 20th century with G. E. Moore’s famous attack upon naturalistic accounts of moral concepts and defense of ethical non-naturalism. According to Moore, theories that try to reconcile morality with naturalism by using natural phenomena to define moral terms are guilty of “the naturalistic fallacy,” which is evidenced by the fact that, whatever natural phenomena these theories pick out as constituting goodness, it remains an open question whether or not those phenomena are themselves good.

The hypothesis that disagreement about the meaning of good is disagreement with regard to the correct analysis of a given whole, may be most plainly seen to be incorrect by consideration of the fact that, whatever definition be offered, it may be always asked, with significance, of the complex so defined, whether it is itself good. And so, on Moore’s account, moral concepts are simple (i.e., have no parts; unexplainable “by any manner of means . . . to anyone who does not already know it”) and indefinable. But, he admits, so are natural properties like being yellow, so why suppose that moral properties are not only unanalyzable but non-natural as well? Moore’s answer involves a further distinction between ultimate and non-ultimate properties. According to Moore, the difference between goodness and simple natural properties like being yellow is that

[good is the only one of these properties whose nature is not at all implicated in other properties. Because higher- and lower-order natural properties differ in their degree of mutual dependency [i.e., dependency upon each other for their explanations], they differ in their degree of ultimacy, with none of them being ultimate. Good, though, is ultimate.]

---

And so moral properties are not only unanalyzable, but non-natural as well.\(^6\)

Also, though he did not actually use the term, Moore is usually credited as the first to suppose that moral properties *supervene* upon natural properties. In his account of intrinsic value, for instance, Moore explains,

> [I]f a given thing possesses any kind of intrinsic value in a certain degree, then not only must the same thing possess it, under all circumstances, in the same degree, but also anything *exactly like it*, must, under all circumstances, possess it in exactly the same degree.\(^7\)

According to Moore, propositions describing such a relation between the moral and non-moral are not only synthetic necessary truths, but also, “the synthetic necessary connections they express are metaphysically rock bottom, and thus are not explainable by any other facts.”\(^8\) All of this, Moore insists, will be obvious to anyone who “will attentively consider with himself what is actually before his mind” when reflecting upon moral concepts, i.e., at least some moral truths are both synthetic and knowable *a priori*.

Moore’s powerful statement of a non-naturalistic moral realism would soon be rejected by logical positivists like A. J. Ayer. Ayer begins his “Critique of Ethics and Theology” by acknowledging a popular interest among philosophers in “the possibility of reducing the whole sphere of ethical terms to non-ethical terms.  We are enquiring whether statements of ethical value can be translated into statements of empirical fact.”\(^9\) After considering two contenders for such a task—utilitarian and subjectivist analyses of moral concepts—Ayer finds them both wanting, for a reason not unlike Moore’s reason for rejecting such theories:

---

\(^6\) There is an interesting body of literature concerning itself with Moore’s meaning of “natural,” and so, what it would mean on Moore’s account for moral properties to be *non*-natural (see, for instance, Moore’s reply to C. D. Broad in Schlipp 1968: 581-592; Hutchinson 2001: 39-60; Copp 2003: 180-181; and Dreier 2006). This matter is ultimately unimportant to the success of this thesis, and so I will refrain from pursuing it entirely.

\(^7\) Moore 1922a: 261.

\(^8\) Horgan 1993: 560.

\(^9\) Ayer 1952: 104.
We cannot agree that to call an action right is to say that of all the actions possible in the circumstances it would cause, or be likely to cause, the greatest happiness, or the greatest balance of pleasure over pain, or the greatest balance of satisfied over unsatisfied desire, because we find that it is not self-contradictory to say that it is sometimes wrong to perform the action which would actually or probably cause the greatest happiness, or the greatest balance of pleasure over pain, or of satisfied over unsatisfied desire. And since it is not self-contradictory to say that some pleasant things are not good, or that some bad things are desired, it cannot be the case that the sentence “$x$ is good” is equivalent to “$x$ is pleasant” or “$x$ is desired.” And to every other variant of utilitarianism with which I am acquainted the same objection can be made.10

So on the unanalyzability of moral concepts, Moore and the positivists agree; but while Moore explained this in terms of the simplicity of moral concepts, Ayer argues, “[T]he reason why they are unanalysable is that they are mere pseudo-concepts. The presence of an ethical symbol in a proposition adds nothing to its factual content.”11 Moral judgments, Ayer is saying, are not truth-responsive to moral facts, but rather are expressions of some empirical fact attended by an emotion of some sort. This, he explains, is because the moral content of such expressions is not empirically verifiable, and so not cognitively significant. For instance, if I were to say, “You acted wrongly in stealing that money,” Ayer explains, it would be “as if I had said, ‘You stole that money,’ in a peculiar tone of horror, or written it with some special exclamation marks.”12 And so, the positivists rejected any cognitivist interpretations of moral judgments, an essential feature of moral realism.13

J. L. Mackie, on the other hand, granted Moore’s cognitivist interpretation of moral judgments, attacking moral realism on the grounds that such judgments are never actually true. Mackie’s two-part argument begins with an observation of “the well-known variation in moral codes from one society to another and from one period to another, and also the differences in moral beliefs between different groups and classes within a complex

---

13 I explain what is “essential” to moral realism in section 1.1 below.
community.” Admitting that it does not follow from this fact that there are no actual moral facts to which moral judgments are truth-responsive, Mackie’s point is just that the actual variations in the moral codes are more readily explained by the hypothesis that they reflect ways of life than by the hypothesis that they express perceptions, most of them seriously inadequate and badly distorted, of objective values.

The second part of Mackie’s argument has two parts of its own—one metaphysical, the other epistemological. On the one hand, if moral properties exist as realists like Moore conceive them, then they are metaphysically “queer” sorts of things—things that are both (a) real, perceivable features of the world and (b) intrinsically motivating.

An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. Similarly, if there were objective principles of right and wrong, any wrong (possible) course of action would have not-to-be-doneness somehow built into it.

Furthermore, Mackie insists, in order to explain our familiarity with these non-natural entities (i.e., moral properties) and the relations they bear to the natural world, the Moorean realist must appeal to some “queer” sort of epistemological faculty: “[A] special sort of intuition’ is a lame answer, but it is the one to which the clear-headed objectivist is compelled to resort.”

So, as of just a few decades ago, moral realism was considered to be a rather taboo position in metaethics. The logical positivists exposed the incompatibility of moral judgments and the standards of empirical verification, while error theorists like Mackie urged that the things we know about the world and human nature are more readily explained by the denial of queer moral entities and faculties of moral perception. Sayre-McCord

---

14 Mackie 1990: 36.
15 Mackie 1990: 37; emphasis mine.
16 Mackie 1990: 40.
summarizes the prevailing sentiment regarding moral realism throughout much of the 20th century thus:

[Ever since Moore offered the Open Question Argument against (definitional) naturalism, philosophers have by and large assumed that moral realism faced insurmountable ontological and epistemological difficulties. Indeed, the common . . . assumption is that the only realist positions available in ethics are those that embrace supernatural properties and special powers of moral intuition.\footnote{Sayre-McCord 1988b: 13.}]

In recent decades, however, these assumptions have been challenged by a number of philosophers interested in developing an account of moral realism squarely within the context of a thoroughgoing naturalism.\footnote{See, for instance, Railton 1986; Boyd 1988; Sturgeon 1988; Brink 1989; and Sayre-McCord 1997.}

In this thesis, I will suggest reasons for thinking that these recent attempts to reconcile moral realism to naturalism have been unsuccessful, for the two are incompatible. In Chapter 2, I will explain how moral properties are most commonly believed to fit into a naturalistic ontology. This will involve a brief discussion of supervenience and the varieties thereof, with particular attention given to the difference between weak and strong conceptions of supervenience. Then I will go on to explain why I think that moral supervenience is problematic for naturalism. Since most (if not all) naturalistic conceptions of moral supervenience describe moral properties as strongly supervening upon the natural world, I will begin by raising a Darwinian dilemma for strong moral supervenience that will push the ethical naturalist into one of two directions: he must either (a) build unique facts and properties about particular social structures (or varying moral belief systems, etc.) into his subvenient sets, and by doing so, jeopardize the objectivity of moral truths that is essential to moral realism, or (b) insist, in the face of all possible worlds (or even different cultures of the actual world) in which people have different moral beliefs than his, that those people are all mistaken—this despite the fact that the belief-forming mechanism responsible for his moral
beliefs was never actually concerned with the truth of those beliefs in the first place. The naturalist still wishing to uphold moral realism will be left wondering whether or not the epistemic freedoms of weak moral supervenience might offer a way out of this dilemma.

In Chapter 3, I will raise a justification problem for weak moral supervenience. The implication will be that, if moral properties only weakly supervene on the natural world, then we are never justified in making moral judgments (and so, it will turn out, weak supervenience is a defeater for moral knowledge). Then I will consider the prospects of some naturalistic version of moral reliabilism as the only means of justifying moral beliefs about weakly supervenient moral facts. My conclusion, however, will be that the prospects are bleak, especially considering the Darwinian problems pointed out in Chapter 2.

In the end, it will look as if neither a strong nor a weak conception of moral supervenience jibes with naturalism. So to conclude, I will suggest that ethical naturalists ought to loosen their grip on something—i.e., they must either accept some version of ethical non-naturalism or be an anti-realist about moral properties. But before I begin, I think it will prove important to make as clear as possible what I will mean when I use the terms “moral realism,” “naturalism,” and “ethical naturalism.”

1.1 What “moral realism” is

Moral realism is generally understood to involve the following two claims: (1) all moral propositions are truth-apt (i.e., they are truth-responsive to the distribution of real moral properties), and (2) at least some moral propositions are true. These two features of moral realism can be seen as in direct conflict with the responses of Ayer and Mackie to Moore described above: (1) is the affirmation of a cognitivist interpretation of moral
judgments, contra Ayer’s expressivism; (2) is the denial of Mackie’s error theory. But more is needed.

Surely, in order to warrant talk of genuine realism, we must understand the instantiation of moral properties (i.e., the content of moral facts) to be a phenomenon entirely independent of the beliefs of any person, group of persons, society, culture, etc. And so some have suggested, and I agree, that moral realism must also be understood as involving something like the following claim: (3) moral facts obtain independently of the thoughts, attitudes, or beliefs of any person, and of the norms, beliefs, or practices of any society or culture. As William Fitzpatrick explains, this is to be understood “in the sense that when it comes to ethics, believing does not make it so, desiring or approving of something does not make it good or right, and so on.” This third condition (which I will call the independence condition) appropriately excludes most, if not all, versions of moral subjectivism and relativism. Fitzpatrick, again, explains:

Ethical reality purchased too cheaply, with ethical truths reduced to truths about what we approve of or practice, for example, would be of too little significance to make for interesting and useful contrasts with non-realist views as such.

Also excluded is any version of constructivism; although accounts like Korsgaard’s do combine cognitivism with a rejection of error theory, constructivism and realism construe moral truths in fundamentally different ways. For the realist, moral truths are things to be discovered, and whatever procedures there are for answering questions in ethics are only correct to the extent that they are responsive to those moral truths. By contrast, the constructivist begins with some procedure that is deemed practically necessary, and so

---

20 See, for instance, Brink 2001: 154; Shafer-Landau 2003: 15; and Fitzpatrick 2006: 2ff. Shafer-Landau expresses this condition in the following way: “[T]here are moral truths that obtain independently of any preferred perspective, in the sense that the moral standards that fix the moral facts are not made true by virtue of their ratification from within any given actual or hypothetical perspective” (2003: 15).


22 Fitzpatrick 2006: 3.
answers to questions in ethics are correct in the constructed sense that they are the result of the given procedure.

1.2 What “naturalism” is

I wish there were more of an agreement among philosophers about how, exactly, naturalism ought to be conceived; but alas, the conceptions are many. Michael Rea, a non-naturalist, insists that naturalism is not properly characterized as a philosophical thesis at all, but rather as a research program, where “research program” is described as “a set of methodological dispositions—dispositions to trust particular cognitive faculties as sources of evidence and to treat particular kinds of experiences and arguments as evidence.”23 Others—naturalists especially—agree, contra Rea, that naturalism is much more than a mere research program; it is a rather important and substantive philosophical view. But still, they disagree about what the substance of that view actually is.

Some seem to regard naturalism as primarily a *metaphysical* thesis, a view about what there is.24 Metaphysical naturalism is roughly the idea that the natural world is all there is, where “natural world” is generally understood to denote something like the causal order of things, or the “spatio-temporal manifold,” as David Armstrong put it.25 Although there is further disagreement among naturalists as to where, exactly, the boundaries of “the natural world” are to be drawn, all seem to agree in their denial of paradigmatically *non*-natural things like souls, angels, God, and the like. There may be a relationship between this metaphysical thesis and Rea’s conception of naturalism as a research program—Rea might suggest that naturalists’ denial of, say, a non-natural God is just the result of them only

23 Rea unpublished MS: 3.
“treat[ing] particular kinds of experiences and arguments as evidence.” But there is an important difference between the two: a research program isn’t in itself the assertion of anything positive, while the thesis of metaphysical naturalism—“The natural world is all there is”—clearly is. In this sense, metaphysical naturalism strikes me as going beyond a mere research program.

Others see naturalism as more of an epistemological thesis, a view about what we can know and how we can know it.\(^\text{26}\) According to David Copp, “A property is natural if and only if any synthetic proposition about its instantiation that can be known, could only be known empirically,”\(^\text{27}\) where “empirical” knowledge is contrasted with its traditional counterpart, \(a priori\) knowledge. Epistemological naturalism, then, is the thesis that all the synthetic propositions that \(can\) be known can only be known empirically. “Ruled out,” Rea observes, are evidential appeals to ungrounded hunches, rational intuitions (conscious episodes in which a proposition seems to be necessarily true), putative divine revelation or religious experiences, manifestly unreliable sources of testimony, and the like.\(^\text{28}\)

This is why naturalists are so uncomfortable with intuitionism. Indeed, the very idea of moral intuitions and their apparent incompatibility with a thoroughgoing naturalism seems to have played a role in driving Ayer and Mackie (and many others) away from moral realism in the first place. (And, as we’ll see, contemporary naturalistic defenses of moral realism construe moral properties as natural things, presumably so that moral propositions can be known empirically.)

There are still other expressions of what the fundamental thesis of naturalism is. Some think that naturalism, at bottom, is a methodological thesis, a view about how philosophical

---

\(^\text{26}\) See, for instance, Quine 1995: 257; Devitt 1998: 45; and Copp 2003.

\(^\text{27}\) Copp 2003: 185.

\(^\text{28}\) Rea unpublished MS: 6-7. The difference between epistemological naturalism and Rea’s idea of naturalism as a research program is less clear than it was in the case of metaphysical naturalism.
Methodological naturalism is the idea that philosophy ought to be done in a way that is continuous with the methods of the natural sciences. But it is hard to see how this is significantly different than Rea’s idea of naturalism as a research program.

Others have added that naturalism ought to be, in some sense, reductive. According to a reductive conception of the natural, there exists a base class of entities that is supposed to be uncontroversially natural, and so for any other entity to be properly considered natural, it must in some way be reducible to a member (or members) of the initial base class. The major difficulty with such an approach is quite simply that many naturalists do not see a need to so constrain our conception of the natural in the first place. It seems perfectly fine, for instance, to suppose that the property being a chair is both (a) irreducible to and (b) just as natural as the physically fundamental things of which the chair is composed. And so there doesn’t seem to be any reason to suppose that non-reductive naturalists are any less naturalistic than those engaged in such a project of reduction.

For my purposes, a thoroughgoing naturalist is someone who affirms both metaphysical and epistemological naturalism, with a mind to ultimately follow the sciences wherever the sciences might lead. So whenever I refer to “naturalism” or “naturalists,” this is what I will have in mind. With Rea, I take it that most naturalists would affirm Wilfrid Sellars’ slogan that “science is the measure of all things: of what is that it is and of what is not.

---

30 Wedgewood (1997) argues that naturalism in ethics just is the reduction of the moral to the natural, and so the only options are (a) reductive naturalism or (b) non-naturalism.
31 Interestingly, Rea (unpublished MS: 5) argues that naturalism as a methodology actually ought to keep people from being either metaphysical or epistemological naturalists. His suggestion is that it is, in principle, possible for the sciences to produce evidence that either the metaphysical or epistemological thesis (or both) is wrong. But, he goes on, since no naturalist would think naturalism itself had been refuted were such evidence to be produced, naturalism must be something different than either of these theses. I don’t think he’s right about this. If the sciences somehow produced compelling-enough evidence for the existence of some non-natural thing (e.g., God) or some non-natural means of knowledge acquisition (e.g., intuitions), my hunch is that any respectable naturalist would just cease being a naturalist. (Indeed, a number of philosophers are coming to embrace various forms of non-naturalism for just these sorts of reasons.) And if I am wrong, I’d suggest that the problem has more to do with the psychology of the unwieldy naturalist than with how naturalism is being conceived.
11

Indeed, this seems to be the very heart of what lies behind all of the various conceptions of naturalism.

1.3 What “ethical naturalism” is

The aim of my thesis is to challenge the adequacy of what many refer to as “ethical naturalism.” Ethical naturalism can mean one of two things. Broadly, ethical naturalism refers to any attempt to reconcile morality to naturalism. In this sense, not only are contemporary moral realists like David Brink and Nicholas Sturgeon ethical naturalists, but so are expressivists like Ayer and error theorists like Mackie. Ayer’s attempt to reconcile morality to naturalism, for instance, led him to be an anti-realist about moral properties. More narrowly, however, ethical naturalism refers to the conjunction of naturalism and moral realism. It is to this narrower (and more common) understanding of ethical naturalism that I will be referring in what follows.

---

CHAPTER TWO: DARWINISM AND STRONG MORAL SUPERVENIENCE

As I pointed out in my introduction, the realism/anti-realism debate in ethics is largely a matter of preserving the importance of morality, on the one hand, without engaging in what Sayre-McCord has called “metaphysical flights of fancy,” on the other. And sure enough, one of the primary reasons that so many have rejected moral realism, especially in the last hundred-or-so years, has been that they find it too metaphysically fanciful. Starting in the 1980’s, however, a number of moral philosophers began to take great strides toward de-mystifying moral realism, explaining how a realm of moral facts and properties can comfortably fit into a strictly naturalistic ontology.

The key for these philosophers has been to explain how it is that moral facts and properties can be conceived of as exhaustively constituted by natural facts and properties. David Brink, for instance, writes,

The ethical naturalist’s appeal to the constitution of moral facts by natural facts is not metaphysically queer. This relation of constitution, composition, or realization is quite familiar. Tables are constituted by certain arrangements of microphysical particles; biological processes such as photosynthesis are constituted by physical and chemical events causally related in certain ways; psychological states are constituted by certain arrangements of brain states; and large scale social events such as wars and elections are constituted by enormously complex sets of smaller scale social and political events causally and temporally related in certain ways.33

Another way of wording this relation of constitution is to say that moral properties supervene upon base sets of natural facts and properties, an idea that, ironically enough, seems to have originated with Moore.34 But unlike Moore, naturalists like Brink take it that moral properties are made up exclusively of the natural things upon which they supervene, making

---

33 Brink 1989: 177.
34 See ch. 1, n. 6.
them no less natural than, say, photosynthesis; and so there is simply no reason for queerness worries.\footnote{There is an interesting debate had by metaphysicians about whether or not the supervenience of some property \textit{A} upon some set of facts and properties \textit{B} commits one to the idea that \textit{A} is something metaphysically distinct from \textit{B}, something “over and above” \textit{B}. When considering, for instance, a metaphysical entity composed of a trout and a turkey, David Lewis writes, “[I]t is neither fish nor fowl, but it is nothing else: it is part fish and part fowl” (Lewis 1991: 80). So the fusion of trout and turkey is metaphysically “nothing over and above” either the trout or the turkey. Others wonder whether or not this makes what sense: “[W]hat does ‘nothing over and above’ mean? This slippery phrase has had a lot of employment in philosophy, but what it means is never explained by its employers” (van Inwagen 1994: 210). Although the debate is both interesting and relevant (Does it ultimately make sense to say that moral properties, while constituted exhaustively by natural things, are “nothing over and above” the things that constitute them?), I will avoid it entirely and assume for the time being that the ethical naturalist’s construal of moral supervenience is “ontologically innocent” enough. For more, see van Inwagen 1990, Merricks 2001, and Dorr and Rosen 2002.}

In this chapter, I wish to explore the idea of the supervenience of the moral realm upon the natural world. In particular, my aim will be to raise worries about the compatibility of moral supervenience with other traditional commitments of naturalism—namely, Darwinism. After spending some time explaining the difference between weak and strong conceptions of supervenience, and why ethical naturalists generally prefer the latter of the two, I will then attempt to show that strong moral supervenience and Darwinian naturalism are incompatible.

### 2.1 Weak and strong supervenience

In his “Concepts of Supervenience,” Jaegwon Kim considers two sets of properties, set \textit{A} and set \textit{B}. Set \textit{A} contains only the property \textit{being a good man}, while set \textit{B} contains the properties \textit{being courageous}, \textit{being benevolent}, and \textit{being honest}. To say that \textit{A} \textit{weakly} supervenes on \textit{B} is to say that if any two men in some possible world have exactly the same \textit{B}-properties in common (say they are both courageous and benevolent but not honest), then the two must either both be good men or neither is. So weak supervenience is the claim:

\[
\text{Necessarily, for any } x \text{ and any } y \text{ in world } \alpha, \text{ if } x \text{ is } A \text{ and } x \text{ is } B \text{ and } A \text{ supervenes upon } B \text{ in } \alpha, \text{ then if } y \text{ is } B, y \text{ will also be } A.
\]
Importantly, on weak supervenience, whether a thing bearing all of the subvening natural properties has or does not have the supervening property as well will depend entirely upon the particular world under consideration, and is not invariant across possible worlds. Weak supervenience, then, allows for all of the following:

(a) In this world anyone who is courageous, benevolent, and honest is a good man, but in another possible world no such man is good; in fact, every such man is evil in this other world.

(b) Again, in this world anyone who has courage, benevolence, and honesty is good; in another world exactly like this one in respect of the distribution of these virtues, no man is good.

(c) In another possible world just like this one in respect of who has, or lacks, these traits of character, every man is good.\textsuperscript{36}

Strong supervenience, on the other hand, fixes the supervenience relation across all possible worlds. Consider again Kim’s two sets of properties. According to strong supervenience, if John is both courageous and honest, and his being a good man \textit{strongly} supervenes upon his being courageous and honest, then anyone in \textit{any} possible world who is also courageous and honest must also be good. And so strong supervenience is the claim:

\[ \text{Necessarily, for any } x \text{ in world } \alpha \text{ and any } y \text{ in world } \beta, \text{ if } x \text{ is } A \text{ and } x \text{ is } B \text{ and } A \text{ supervenes upon } B \text{ in } \alpha, \text{ then if } y \text{ is } B, y \text{ will also be } A. \]

If moral properties strongly supervene on natural properties, then there cannot be two naturally indiscernible worlds that are morally discernible because the relationship between the natural world and the realm of moral facts and properties is a necessary one.

So when philosophers say that moral facts and properties supervene upon natural facts and properties, to which sort of supervenience are they referring, weak or strong? R. M. Hare once said that his claim that value properties supervene on other sorts of properties was intended to be an appeal to weak supervenience.\textsuperscript{37} But among moral philosophers, Hare is very much in the minority, for reasons often associated with the work of Simon

\textsuperscript{36} Kim 1984: 159.

\textsuperscript{37} Hare 1984: 4.
According to Blackburn, philosophers who suppose that moral facts and properties only weakly supervene upon natural facts and properties cannot explain why there cannot possibly be two things in the same world that are indiscernible with respect to their subvening natural properties but discernible with respect to their supervening moral properties. For these and other reasons, ethical naturalists generally tend to favor a strong conception of moral supervenience.

2.2 Dependence and relevance

There are two more things that I'd like to point out about moral supervenience that will prove important for the purposes of this thesis. First, moral supervenience can be understood as a kind of dependence relation, according to which a thing has the moral property that it has because it has the natural properties that it has. In other words, a thing’s moral value will depend upon its natural characteristics. In this sense, borrowing Dancy’s term, moral properties are resultant properties: They are the metaphysical result of the coming together of certain natural facts and properties.

Second, a distinction can be made between relevant and irrelevant properties when enumerating the members of our subvenient sets. If moral properties are resultant properties that depend for their presence and content upon the natural features that subvene under them, then subvenient sets will include only those things that are relevant to the determination of supervening properties. Consider the act of rape. Presumably, some

---

39 Blackburn 1985: 134ff. Blackburn calls such a world a “mixed world.”
40 I should note that, for the purposes of my argument, the question of whether moral properties supervene weakly or strongly will turn out to be unimportant, since I will argue that both are troublesome.
41 Although there are some supervenience relations that cannot be characterized in such terms (e.g., for any property F, being F will supervene on being not-F, since two things cannot differ with respect to being F without also differing with respect to being not-F; but nothing is F because it is not-F), I see no problem so characterizing the supervenience of the moral upon the natural.
42 Dancy 1993: 73.
natural, descriptive properties of that event subvene under the moral property being wrong, while other properties of the event do not. The property being the cause of suffering, for instance, probably plays a significant role in the instantiation of wrongness. But other properties—like, say, taking place near a tree or taking place three weeks before Flag Day—while certainly parts of a full description of the event, are not constitutive of moral wrongness.

This is what Blackburn has in mind with his “limitation thesis”:

This [limitation thesis] will say that whenever a property $F$ supervenes upon some basis, there is necessarily a boundary to the kind of $G$ properties that it can depend upon. . . . [T]he moral supervenes upon the natural, and the thesis will tell us that there are some natural properties that necessarily have no relevance to moral ones—pure spatial position, perhaps, or date of beginning in time.43

So when we are asked to name some, or all, of the things that subvene under a moral fact or property, we are only concerned with those things that play a relevant, morally-determinant role.

2.3 Darwinism and ethical naturalism

As per my description of naturalism in Chapter 1, I will assume for the sake of this discussion that ethical naturalists want to uphold the possibility of a scientific explanation of the world that we inhabit. Naturalism, I remarked, is understood by many as a kind of project (or substantive philosophical thesis) according to which the world is explainable in terms of the subject matter of the various scientific disciplines.44 The function of white blood cells, the nature of human emotions, and the rituals of some remote Indonesian tribe are not typically understood by naturalists to be best described in terms of, say, the supernatural, but rather as the subject matter of the biological, psychological, and anthropological sciences, respectively.

43 Blackburn 1985: 133.
A further assumption that I think it probably safe to make is this: the various full accounts that naturalists will give for such things as white blood cells, for instance, will all involve a harkening back to the *evolutionary* origins of those things. This is because, as Sommers and Rosenberg observe, “Darwinian naturalism is the only game in town for naturalists.” So if, as Daniel Dennett has argued, Darwinism is indeed a kind of “universal acid” that eats through and informs our understanding of just about everything, then ethics cannot be the sort of autonomous field of enquiry that others think it is. After all, as long as naturalism is understood as the kind of far-reaching explanatory project that it is, how odd it would be to appeal to evolution for explanations of the nature and origin of white blood cells and the human emotions, for instance, but to then offer some *other* sort of account of the nature and origin of human morality. Surely, for the ethical naturalist, the nearest available account of the origin of our beliefs about morality, and the account that squares most comfortably with his or her naturalism, is an evolutionary one. But, as I will now explain, despite the fact that most ethical naturalists regard it a *virtue* of their theory that it squares so nicely with the explanatory aims of the evolutionary sciences, the story that evolution actually tells about the formation of our moral beliefs is one that seems, upon inspection, to be incompatible with the implications and demands of strong moral supervenience.

### 2.4 Evolution and the formation of our moral beliefs

So what does evolution have to say about our having the moral beliefs that we have? On a Darwinian scheme, we have the moral beliefs that we have, to a significant degree at

---

45 Sommers and Rosenberg 2003: 659.
least, because of selective evolutionary pressures. We should expect there to have been tremendous evolutionary pressure to hold, for instance, those moral beliefs that tend more so than others to promote survival and reproductive success. But if evolutionary pressures have played such a role in the formation of our moral beliefs, pushing us in the direction of some rather than others, then what reason is there to suppose that we’ve arrived upon a body of true moral beliefs? So the question for the ethical naturalist, raised recently by Sharon Street,48 becomes this: Given these potentially distorting evolutionary pressures, what sort of relationship is there between our having the moral beliefs that we have, on the one hand, and the truth-makers of those beliefs (i.e., the actual distribution of moral properties), on the other?

Street considers two competing accounts of this relationship—the tracking account and the adaptive link account.49 According to the tracking account, although the content of our moral beliefs has been largely determined by evolutionary pressures, “we may understand these evolutionary causes as having tracked the truth.”50 Our having these moral beliefs is explained by the fact that the beliefs themselves are true, “and that the capacity to discern such truths proved advantageous for the purposes of survival and reproduction.”51 On the other hand, according to the adaptive link account, having certain moral beliefs rather than others contributed to the reproductive fitness of our ancestors not because the beliefs are true, but simply

48 Street 2006.
49 Street 2006: 125-135. Street does not actually consider these to be accounts of the relationship between our having the moral beliefs that we have and the truth-makers of those beliefs, but rather accounts of the relationship between “the selective forces that have influenced the content of our evaluative judgments . . . and the independent evaluative truths that realism posits” (2006: 1). But, as she discusses them, they are just as easily seen as accounts of how the moral beliefs that have arisen due to particular evolutionary forces are related to the truth-makers of those beliefs. She is, in her paper, making roughly the same point that I wish to make in this section of mine.
50 Street 2006: 125.
51 Street 2006: 126.
because they forged adaptive links between our ancestors’ circumstances and their responses to those circumstances, getting them to act, feel, and believe in ways that turned out to be reproductively advantageous.\textsuperscript{52}

So whereas on the tracking account, there is a dependence relationship between our having the moral beliefs we have and the truth-makers of those beliefs—with the former depending upon the latter for their presence and content—there is no such dependence according to the adaptive link account.

Street presses her point further by suggesting that there are good reasons to think that the adaptive link account is explanatorily superior to the tracking account. First of all, while both may offer compelling stories about how we’ve come to have the moral beliefs that we have, “the tracking account posits something extra that the adaptive link account does not, namely independent evaluative truths.”\textsuperscript{53} The adaptive link account is therefore the more parsimonious of the two. Second, the adaptive link account is in some sense clearer than the tracking account, since it is not clear how having certain moral beliefs \textit{because they are true} would contribute to our survival and reproductive success any more so than having the same moral beliefs for other, obviously helpful reasons (e.g., their tendency to forge adaptive links necessary for survival). And finally, the adaptive link account does a better job of explaining “why there are widespread tendencies among human beings to make some evaluative judgments rather than others.”\textsuperscript{54} Not only does the tracking account fail to explain the remarkable coincidence that so many moral truths turn out to be just the sorts of things that forge the adaptive links that Street describes, but it also fails to explain our natural tendency to hold some moral beliefs that are probably \textit{not} true (e.g., the belief that

\textsuperscript{52} Street 2006: 127.
\textsuperscript{53} Street 2006: 129.
\textsuperscript{54} Street 2006: 132.
“the fact that someone is in an ‘out-group’ of some kind is a reason to accord him or her lesser treatment than those in the ‘in-group’.”

What the ethical naturalist should find especially troubling about all of this is precisely that, on the better of the two evolutionary explanations of our having the moral beliefs that we have, the truth of those beliefs seems to play no part whatsoever in the formation of those beliefs. According to Sommers and Rosenberg, the Darwinian account of the origins of our moral beliefs makes it “apparent that the best explanation—blind variation and natural selection—for the emergence of our ethical belief[s] does not require that these beliefs have truth-makers.” In the next section, I will explain the implications that I think this conclusion has upon the ethical naturalist’s claim of strong moral supervenience.

2.5 A Darwinian dilemma for strong moral supervenience: the first horn

Given the many contingencies of evolution, it seems reasonable to suppose that the human social structures upon which most (or all) of our moral beliefs are to some extent based could have been significantly different. Darwin himself imagined a possible world in which human beings developed under social conditions not unlike those of hive-bees. In such a world, it is reasonable to suppose that circumstances could obtain in which all of the relevant natural properties are present, and upon which, in our world, we believe wrongness supervenes, but nonetheless, in that world, people believe that some property other than wrongness supervenes (or perhaps people believe that no moral property supervenes at all). This much is possible even if moral properties strongly supervene on the natural world.

---

55 Street 2006: 133.
56 One might wonder why we can’t reflect upon the content of our moral beliefs and critically evaluate them in a more truth-minded sort of way. I address this sort of objection below.
57 Sommers and Rosenberg 2003: 667.
58 Darwin 1981: 73. Darwin, of course, did not use the term “possible world.”
Were we like hive-bees, for instance, we would believe our unmarried females to have some duty to kill their brothers. And so, whereas we would (hopefully) believe my sister’s killing me in the actual world to be *wrong*, there is a not-too-distant possible world (call it W) in which our beliefs about the matter are quite different, despite all of the *relevant* natural properties obtaining.

Call the set of relevant natural properties involved in my sister’s killing me in the actual world $K_1$, and we’re supposing that wrongness supervenes on $K_1$. In W, however, people are like hive-bees, and so a relevantly identical set of natural properties, $K_2$, is *believed* to yield, say, no moral property at all. Now, according to strong supervenience, to imagine a circumstance different with respect to its supervening property must be to imagine it different with respect to its subvening properties as well, no matter the possible world. So, if it is the case that moral properties strongly supervene on natural properties, then the ethical naturalist has two options with regard to this thought experiment: either (a) *deny* that $K_1$ and $K_2$ are identical in every relevant respect, or (b) agree that the two circumstances really *are* identical with respect to their moral properties, which would mean that one group of people—those in the actual world or those in W—is actually quite mistaken in its moral beliefs about siblicide (and, presumably, many other moral beliefs). But, as I will explain, neither (a) nor (b) is in the end appealing.

Consider the first of the ethical naturalist’s two options and recall the distinction I made above between relevant and irrelevant properties. The question is not, “How might $K_1$ and $K_2$ be naturally discernible?” but rather, “How might $K_1$ and $K_2$ be *relevantly* naturally discernible?” While some might answer that the fact that $K_2$ occurred in a world where people have evolved socially like hive-bees, having hive-bee sorts of beliefs and patterns of behavior, etc., these sorts of considerations cannot be counted as relevant by moral realists.
As I explained earlier, according to moral realism, moral facts cannot in any way depend upon the beliefs of any person or norms of any society. And so such facts cannot play a determinant role in the moral nature of any person, action, or state of affairs. In other words, despite the fact that people in W believe very sincerely that my sister has some sort of sacred duty to kill me, the independence condition of moral realism makes it such that “believing does not make it so, desiring or approving of something does not make it good or right, and so on.”

After all, if considerations about either the beliefs of persons or facts about individual social structures are allowed to be relevant in our evaluation of \( K_1 \) and \( K_2 \), then I see no reason to disallow these sorts of considerations to be relevant when comparing similar circumstances occurring within different social structures in the same world. There are, of course, cultures in the actual world that exhibit vastly different moral beliefs and practices. In their book *Unto Others*, Elliot Sober and David Wilson mention a recent study that found that 39 of 60 randomly selected societies practiced infanticide.\(^{59}\) So if we’re building information about different social structures into our subvenient sets of relevant facts and properties, then perhaps wrongness does not (in certain cultures) supervene upon the killing of a female infant as long as the people doing so believe, say, that male offspring are of greater value than females. And perhaps, due to allegedly relevant differences in social structure or environment, the “ironing” of a young woman’s breasts in order to make her less attractive to on-looking men, despite (presumably) subvening under wrongness here in egalitarian America, does not subvene under wrongness in Cameroon (or in parts of Cameroon, or for certain inhabitants of Cameroon).\(^{60}\) I could go on and on, but this is just relativism—something any realist ought to avoid—and it has everything to do with allowing

---


\(^{60}\) Sa’ah 2006.
too much into our subvenient sets of relevant, morally-determinant natural facts and properties. So the challenge for the ethical naturalist will be to suggest some relevant natural difference between circumstances like \( K_1 \) and \( K_2 \) that does not, again borrowing Fitzpatrick’s words, purchase ethical reality too cheaply, “with ethical truths reduced to truths about what we approve of or practice.”

2.5.1 Defending the first horn

The reader might ask, “But what if, due to relevant differences of environment and circumstance, the divergent moral beliefs or practices of a particular culture are justified?” After all, it certainly seems possible for there to be naturalistically respectable reasons for some culture to believe infanticide to be permissible (e.g., population pressures). So why not suppose that whatever it is that justifies such beliefs or practices (e.g., infanticide’s survival value) is itself a relevant natural difference between two cultures and the beliefs thereof?

My response is that while the naturalist is fully capable of providing very robust and informative accounts of what is and is not necessary for a species’ (or culture’s, or person’s) survival, we’ve got reason to think, via Street’s argument, that this is not only not the same thing as the apprehension of moral facts, but also wholly unrelated to the obtaining of moral facts. Street’s argument drives a kind of epistemic wedge between facts about survival value and facts about moral right and wrong, and ethical naturalists seem to be just begging the question against such an argument by assuming that the two sorts of facts are one and the same, or that they’re even related. I take it that there is an important difference between propositions like “Exercising is good if you want to prolong your life” or “Driving with your eyes closed is bad if you want to prolong your life” and propositions like “Saving lives is good” or “Taking lives is bad.” The first two propositions seem to be ascriptions of some non-
moral value to an action, whereas the last two seem clearly to be moral propositions (I have in mind something like the difference between instrumental and moral good). This distinction bears an important relationship to Street’s distinction between the two accounts of our moral beliefs, the truth-tracking account and the adaptive-link account. What is suggested by the adaptive-link account is precisely that whenever we think we’ve got our hands on moral truth, what we’ve really got our hands on is just an inherited belief about non-moral survival value. So the question that the ethical naturalist will now have to answer is this: Why suppose that the fact that some culture’s belief or practice is justified (a better word might be demanded) by certain features of its circumstances is ever more than just a natural difference, but rather a relevant, morally-determinant natural difference? And the naturalist’s answer will have to be something more interesting than merely defining moral concepts in terms of that which is needed for survival. Not only would such a response run the risk of conflating instrumental and moral goods, but it would also have to provide a principle (or system of principles) for weighing certain instrumental goods against each other whenever conflict between two or more arises, and this would seem to invoke a concept of normativity that goes beyond the reach of naturalism.  

2.6 The second horn of the dilemma

What about the ethical naturalist’s second way out of the dilemma? Here she must insist that that someone is mistaken. If wrongness does supervene on $K_1$, then it supervenes on $K_2$ as well, and people in $W$ are just wrong to believe otherwise. On the other hand, if people in $W$ are right, and wrongness does not supervene on $K_2$, then it is those of us in the actual world that are mistaken in thinking that $K_1$ is wrong. After all, “wrongness” is either

---

61 I.e., is there a strictly naturalistic reason for counting certain goods as more important than others? And “more important” in what sense?
constituted by that set of relevant natural properties or it is not; just like “water” either consists in $\text{H}_2\text{O}$ or it does not. So for those of us in the actual world to say of K-actions that they are wrong, we must then be implying that people in W are mistaken in their moral judgments. But, first of all, by implying that people in W are so mistaken, aren’t we thereby guilty of what Judith Jarvis Thompson has called “metaphysical imperialism,” where we wrongly just assume that the meaning of normative terms is inextricably tied to our interests, our language, and/or our social structure? And second, where do we get off doing such a thing if, on an evolutionary scheme, we’ve got no compelling reason to think that the mechanism responsible for our moral beliefs was ever truth-aimed in the first place? It certainly seems as if a Darwinian account of the formation of our moral beliefs might here serve as a kind of defeater to any claims of moral authority that people in the actual world (or any world with a similar Darwinian story to tell) might be able to muster.

2.6.1 Defending the second horn

In response to this suggestion that Darwinian considerations give us reason for moral skepticism, and so remove any grounds of moral authority upon which we might judge the moral beliefs of others to be right or wrong, the ethical naturalist might ask if the trustworthiness of our moral beliefs might be redeemed by an appeal to our capacity for rational reflection. Surely, the objector might suggest, we are able to reflect on the moral beliefs that we have and critically evaluate them in a way that might “latch onto” moral truth.

---

62 Harman and Thompson 1996: 136. Thompson’s example is that of a hammer’s being a “good” hammer. Presumably, she writes, “there are a lot of odd possible worlds, and I am sure there are some in which the wants people typically hammer in nails to satisfy are very different” (1996: 135). Only a metaphysical imperialist, Thompson suggests, would “insist that the people in that world hammer nails in for all the wrong reasons, and that large slabs of granite are not in any world good for use in hammering in nails” (1996: 136). Her point is closely associated to the point Horgan and Timmons make about “tether[ing] the meaning and reference of the relevant terms too tightly to a theory that has some special connection to human beings” (see Horgan and Timmons 1996). The call it “conceptual chauvinism.”
Here is Street’s response to this sort of objection:

[What rational reflection about evaluative matters involves, inescapably, is assessing some evaluative judgments in terms of others. . . . The widespread consensus that the method of reflective equilibrium, broadly understood, is our sole means of proceeding in ethics is an acknowledgement of this fact: ultimately, we can test our evaluative judgments only by testing their consistency with our other evaluative judgments, combined of course with judgments about the (non-evaluative) facts. Thus, if the fund of evaluative judgments with which human reflection began was thoroughly contaminated with illegitimate influence [i.e., evolutionary pressures unconcerned with truth] . . . then the tools of rational reflection were equally contaminated, for the latter are always just a subset of the former. It follows that all our reflection over the ages has really just been a process of assessing evaluative judgments that are mostly off the mark in terms of others that are mostly off the mark.]

Another way of putting Street’s point is to ask: Why suppose that the processes of critical reflection to which the objector refers are ever reliable when it comes to getting our hands on moral truths? Epistemic reliability has everything to do with a mechanism’s (or process’s) having produced more true beliefs than false beliefs. But if the evolutionary considerations I’ve described above are enough for even a temporary skepticism of the truth-responsiveness of our moral beliefs, then upon what grounds might we be able to make any claims whatsoever about reliability? In other words, some critically reflective process might very well land us upon some moral truth of the matter, but (a) we’d never know that it bad, and so (b) we’d have no basis for making claims about what sort of process is genuinely reliable and what processes aren’t.

2.7 Conclusion

Ethical naturalists may now have reason to worry that strong moral supervenience does not square as comfortably with other commitments of theirs (i.e., Darwinism) as they had once thought. So inasmuch as they are interested in upholding ethical naturalism, they might wonder if weakening the supervenience claim might offer them a solution to my dilemma. After all, if the moral realm only weakly supervenes upon natural facts and

---

63 Street 124.
64 I offer a more thorough evaluation of the prospects of a naturalistic moral reliabilism in Chapter 3.
properties, then there is no problem at all with $K_1$ and $K_2$ subvening under different moral properties, and so the above dilemma is easily answerable. Unfortunately, as I will argue in Chapter 3, weak supervenience is a defeater for moral knowledge, and so not the sort of problem-solver that an ethical naturalist might hope.
CHAPTER THREE: WEAK MORAL SUPERVENIENCE AND JUSTIFICATION

As I explained earlier, on weak supervenience, there is a possible world identical to ours in every natural respect, but different in terms of its configuration of moral properties across those natural phenomena. In fact, if moral properties only weakly supervene on natural properties, then there is a possible world that is identical to ours in every natural respect but contains no moral properties at all. Call this world W*.

In W*, as opposed to W in Chapter 2, evolutionary processes have taken precisely the same course that they have in our world, resulting in precisely the same cultural phenomena, social norms and practices, and other sorts of natural facts, whether significant to my discussion or not. So in W*, a man exactly like the Adolf Hitler of the actual world in every natural respect led exactly the same life, in every natural respect. Call him Schmadolf Hitler. However, the sort of moral character that we typically ascribe to the Hitler of our world (e.g., morally depraved), and the sorts of moral properties we believe his actions to have had (e.g., evil), are not actually a part of the ontology of W*. So in W*, Schmadolf Hitler is also causally responsible for the termination of millions of people, and when people in W* reflect upon his actions, they experience the same sort of emotional uneasiness that people in our world experience when they reflect upon the events of the Holocaust. But when we say that the Holocaust was evil, or that Hitler was depraved, those terms actually have metaphysical referents in our world (i.e., moral properties), whereas such words bear no genuine connection to reality in W*, for there are no such properties in that world. The questions I want to explore in this chapter are these: Are we justified in supposing there to

---

65 Or perhaps Adolf Schitler, if you prefer. I should note that by “every natural respect,” I mean not only that Schmadolf and Adolf are exactly similar biologically, physiologically, etc., but also that even the events of their lives up until this point have been exactly alike, all except for any sorts of moral truths about those events.
be any real difference between W* and the actual world? And if not, then in what sense are we justified in believing there to be moral properties in the actual world?

As we’ve already established, W* is a world identical to ours in every natural respect, differing only in the sense that there are no moral properties in W*. But if the two worlds are naturally indiscernible, then so must whatever faculties are responsible for our belief in the existence of supervening moral properties in our world be shared by people in W*. So inhabitants of W* have precisely the same cognitive faculties performing precisely the same relevant functions, arriving at precisely the same moral beliefs. Their neuropsychological makeup is no different than ours. As I mentioned earlier, people in W* have the same psychological response to events in their world as we do to events in ours—e.g., repulsion at the thought of an event like the Holocaust, leading them to regard it as evil. If, however, a person in W* were to say of some action that it is wrong, or of some state of affairs that it is evil, she would only be fooling herself because there is no such thing in W*. But if this is allowed, then upon what grounds could we possibly claim to know that our world actually does have these supervening moral properties while W* does not? If inhabitants of W* can have precisely the same sorts and patterns of (what might be called) “moral” experiences as we do and still be mistaken about the reality of moral properties, then so could we. It would seem, then, until we are able to distinguish the actual world from W*, there is some reason to doubt that our moral propositions are ever true, for it might be the case, like it is for those in W*, that our moral terms actually refer to nothing. And so, perhaps some variety of moral skepticism might be more appropriate, that is, if we assume that moral properties only weakly supervene on natural properties.

At this point, I suspect someone will wonder, “How is this problem (for weak moral supervenience) any more threatening to the security of our moral knowledge than brain-in-a-
vat/Cartesian deceiver stories are to the security of our knowledge in general?" As long as moral properties do supervene on natural properties in the actual world, and as long as we’ve got some positive reason to suppose that moral properties supervene on natural properties in the actual world (e.g., perhaps because they figure into our best explanations of various phenomena), then why suppose that we’ve got to know that moral properties supervene on natural properties in the actual world in order to have moral knowledge—that is, why suppose that we have to know that we know? In what follows, I will argue that, given weak supervenience, naturalists do not have positive reason to suppose that moral properties supervene on natural phenomena in the actual world. So if my argument is sound, it will present those ethical naturalists settling for the freedoms of weak supervenience with a defeater for moral knowledge.

3.1 Weak supervenience and the justification of moral beliefs

In his article “Moral Explanations,” Nicholas Sturgeon makes a case for the explanatory potency of moral properties. In doing so, Sturgeon recalls Bernard DeVoto’s *The Year of Decision: 1846*, which tells the story of a group of American emigrants already in California and their efforts “to rescue another party of emigrants, the Donner Party, trapped by snows in the High Sierras.”66 The rescue efforts, however, were lead by a certain Selim Woodworth, who “not only failed to lead rescue parties into the mountains himself, . . . but had to be ‘shamed, threatened, and bullied’ even into organizing the efforts of others willing to take the risk.”67 In his free time, Woodworth was sure to arrange for his own comforts in camp, ironically touting the importance of his position. And, of course, many innocent emigrants died who might have been spared, “including four known still to be alive when he

---

66 Sturgeon 1988: 244.
[Woodworth] turned back for the last time in mid-March.\textsuperscript{68} Thus, DeVoto concludes, “Woodworth was just no damned good.”\textsuperscript{69}

Getting back to his case for the explanatory value of moral facts and properties, Sturgeon urges, “If DeVoto is right about this evidence, . . . it seems reasonable that part of the explanation of his believing that Woodworth was no damned good is just that Woodworth \textit{was} no damned good.”\textsuperscript{70} If moral facts (e.g., Woodworth’s \textit{being no damned good}) are explanatorily irrelevant to the circumstances in which they are believed to obtain, a suggestion made by Gilbert Harman,\textsuperscript{71} then it seems to Sturgeon reasonable to suppose not only that the circumstances would have occurred in just the same way, but also that we could just as easily (and accurately) explain those circumstances, even if the moral facts had been otherwise.

For it is natural to think that if a particular assumption is completely irrelevant to the explanation of a certain fact, then the fact would have obtained, and we could have explained it just as well, even if the assumption had been false.\textsuperscript{72}

But in what reasonable way might the above story be told, preserving all of its non-moral characteristics, if Woodworth had been, say, \textit{so} damned good? Does it seem reasonable to suppose that Woodworth would have done the things he did if he had been \textit{so damned good}? Presumably not. Woodworth did the things he did, at least in part, because of his despicable moral character. So it would seem that, in some sense, the moral realm has a determinant causal power over the natural facts of human activity.

But surely this does not square well with the allowances of weak supervenience. After all, as difficult as it might be for us to imagine Selim Woodworth or, say, Adolf Hitler having done the things they did without also having had the sort of depraved moral

\textsuperscript{68} Sturgeon 1988: 244.
\textsuperscript{69} DeVoto 1942: 442.
\textsuperscript{70} Sturgeon 1988: 244.
\textsuperscript{71} Harman 1977: 3-10.
\textsuperscript{72} Sturgeon 1988: 245.
characters that we believe they had, it is, by definition, possible given weak supervenience—witness Schmadolf Hitler. And further, if people in W* have precisely the same sorts of beliefs about Schmadolf as we have about Adolf, then in what sense can we really insist that Adolf’s actual moral depravity played some role in the formation of our belief that he was morally depraved? After all, any account we give is only going to be echoed by people in W*, but, according to our thought experiment, they’d be exactly wrong. So on weak supervenience, not only do we have no way of knowing whether or not moral properties really do supervene on natural properties in the actual world (i.e., whether or not the actual world and W* are morally discernible), but we may also lack justification for our moral beliefs, since any justificatory accounts of ours could be (mistakenly) echoed by people in W*. And so weak moral supervenience is a defeater for moral knowledge.

To be clear, the suggestion I am making is this: If the only difference between W* and the actual world is that the actual world contains moral properties and W* does not, and there is no justified way to distinguish the actual world from W*, then we cannot justifiably believe that there are moral properties in the actual world. Consider a similar case. After a game of basketball with a few friends, I walk into the locker room to grab my gym bag before leaving. At the same time, a friend of mine returns to the locker to grab his gym bag. While we were playing, a janitor tossed all of our bags into the same corner of the locker room, and so they are not in the places we remember leaving them. Suppose my friend and I have qualitatively identical bags, filled with qualitatively identical items. The only difference between the two bags is that one belongs to me and the other belongs to him, but there is no way for either of us to distinguish between the two. In such a case, I hardly seem justified grabbing one and believing that it is mine. So the same principle that I am suggesting applies to W* and the actual world seems to apply just as well to the two
identical gym bags: If the only difference between two gym bags is that one belongs to me and the other belongs to my friend, and there is no justified way to distinguish one bag from the other, then I cannot justifiably grab one and believe that it is mine.

Notice that it will not help at all to point to any particular feature of one of the bags—say, for instance, to point to a small tear near one of the ends of the zipper—because the two bags are identical. In response to any such suggestion, my friend will only have to say, "But *that* bag has the same-sized tear in the same place." And so the fact that the bag I'm holding has a small tear near one of the ends of the zipper cannot count as a justifying reason for believing that "*This* bag is mine," since there will be a counter-weighing defeater immediately available for every such reason. Similarly, when trying to distinguish between the actual world and \( W^* \), it will not help to point to any of the features of the actual world that are traditionally taken to be justifying reasons for moral beliefs (e.g., the near universal disapproval of Hitler's behavior), since there will be a counter-weighing defeater immediately available in \( W^* \) for every such reason.

### 3.2 Alternative hypotheses arguments

In his "Moral Skepticism and Justification," Walter Sinnott-Armstrong develops two skeptical arguments against the justification of moral beliefs.\(^{73}\) According to one, which I will call the alternative hypotheses argument, no person is ever justified in holding any moral belief unless he or she is able to rule out alternative hypotheses. And so, as it turns out, no person is ever justified in holding any moral belief since no person can ever rule out the possibility that moral nihilism is true (without begging the question, that is). The argument that I am making here very closely resembles an alternative hypotheses argument. I am

---

\(^{73}\) Sinnott-Armstrong 1996: 8-17.
saying that naturalists cannot be justified in believing the actual world to contain moral properties unless they can rule out the alternative hypothesis that $W^*$ and the actual world are identical. And since, as I'm arguing, naturalists cannot rule out this alternative hypothesis, they are not justified in believing the actual world to contain moral properties (and, consequently, they are not justified in any particular moral belief about the actual world).

There are generally two ways of responding to alternative hypotheses arguments. According to the first, one tries to distinguish between likely and unlikely alternative hypotheses and adds that only likely alternative hypotheses are genuinely threatening. To borrow one of Sinnott-Armstrong's examples, imagine Hannah is trying to figure what is inside a wrapped birthday present. The present is shaped like a small flat box and its contents rattle when she shakes it. Since it’s unlikely that the present contains either a bike (because the box is too small) or clothing (because clothing doesn’t often rattle when shaken), she can justifiably rule out these alternative hypotheses. But since it seems (equally) likely that the present contains either a puzzle or a game, she cannot justifiably rule out either of these alternative hypotheses, and so she cannot justifiably believe that the present contains a puzzle.

This strategy will not work with the cases I have been considering. Here are two hypotheses:

1. *This* gym bag is mine

and

2. *That* gym bag is mine.

Each of these hypotheses, given the circumstances of the case as I have described them, is equally likely to be true. While Hannah seems to be justified in believing, “There is no bike
in this box because this box is too small,” there is no similarly justifiable way for me to finish this sentence: “*This* gym bag is mine because . . .” Here are two more hypotheses:

3. The actual world contains moral properties but $W^*$ does not

and

4. Neither the actual world nor $W^*$ contains moral properties (i.e., they are identical).

Again, each of these hypotheses, given the circumstances of the case as I have described them, is equally likely to be true. And so the distinction between likely and unlikely alternative hypotheses seems unhelpful as a strategy for defending the justification of moral beliefs given weak moral supervenience.

Alternative hypotheses arguments in epistemology are also dealt with by appealing to the non-necessity of *certainty* for knowledge. “Sure,” one might respond, “it’s *possible* that all of my beliefs about the external world are the byproducts of some evil demon’s deceptive schemes, and so I cannot be *certain* that any of those beliefs are true, but why suppose that I need such certainty in order to have genuine knowledge of the external world?” I happen to find responses like this rather compelling, and so I’ve never thought too much of alternative hypotheses arguments in epistemology. But can’t something similar be said in response to the problem that I’ve been raising for justifying moral beliefs given weak supervenience? Can’t the ethical naturalist just say something like, “Sure, we might not be able to rule out the hypothesis that $W^*$ and the actual world are identical (in that neither contain moral properties), and so we cannot be certain that moral propositions are ever true in the actual world, but that sort of certainty is not necessary for moral knowledge”?

The answer to this question is yes, the naturalist is free to suggest that moral certainty is unnecessary, and so we might still be justified in believing the actual world to contain moral properties even if we are unable to rule out hypothesis (4) above; but more
will be needed. After all, there is still this question: How might we be justified in believing the actual world to contain moral properties, or in holding any moral belief in particular, despite our inability to rule out hypothesis (4)? As I see it, there are two answers that might be given here; but, I will argue, neither of them is available to the ethical naturalist.

The first answer, given by Russ Shafer-Landau, is simply to admit the possibility of self-evident moral beliefs. He explains,

Beliefs are self-evident if they have as their content self-evident propositions. A proposition p is self-evident = df. p is such that adequately understanding and attentively considering just p is sufficient to justify believing that p.\(^{74}\)

If p is self-evident, then one’s justification for believing that p is not conditioned upon one’s ability to eliminate alternative hypotheses, and so the admission of self-evidence amounts to the denial of the central premise of alternative hypotheses arguments—that one is justified only to the extent that he or she is able to eliminate alternative hypotheses. As candidate self-evident moral beliefs, Shafer-Landau suggests the following: Other things being equal, (a) it is wrong to take pleasure in another’s pain, (b) it is wrong to taunt and threaten the vulnerable, (c) it is wrong to prosecute and punish those known to be innocent, and (d) it is wrong to sell another’s secrets solely for personal gain. And surely, these do strike many as obviously true. But Shafer-Landau is only free to make such appeals to self-evidence because he’s a non-naturalist. As defined in my introduction, a thoroughgoing naturalism is committed not just to metaphysical naturalism (which, I gather,Shafer-Landau affirms\(^{75}\)) but also epistemological naturalism (which, with his appeals to self-evidence and his denial that moral knowledge is always an empirical matter, Shafer-Landau denies). And so, whether successful or not, self-evidence is not an answer that is available to ethical naturalists.

---

\(^{74}\) Shafer-Landau 2003: 247.

\(^{75}\) While Shafer-Landau affirms that moral properties and the natural properties that subvene under them are metaphysically distinct (and so, he denies reductivism in ethics), he does insist that “moral properties are constituted exclusively by instantiations of descriptive properties” (2003: 78; emphasis mine). And so moral properties are not a different sort of stuff, metaphysically speaking.
3.3 Moral reliabilism

The second way to explain how we might be justified in our moral beliefs despite our inability to rule out hypothesis (4), and one that is not so clearly unavailable to ethical naturalists, will involve an appeal to some externalist account of moral justification. After all, the problem is essentially that none of the resources available internally to us are enough to distinguish the actual world from $W^*$, and so the justification of our moral beliefs will have to be an external matter. Consider two types of justification:

The first asks of a given believer whether she is justified in holding a belief. Call this agent justification. The second asks of a given believer whether she is able to show that her belief is plausible. Call this demonstrative justification.\(^{76}\)

People often fail to distinguish between these two sorts of justification, and so it is often assumed that a person is only ever justified if she is demonstratively justified. But after considering three ways of conceiving the standards for demonstrative justification and finding all of them problematic, Shafer-Landau suggests that this assumption is just wrong.

According to the first of these conceptions, one justifies one’s belief if and only if one is able to produce the reasons that support it. But surely believing certain mathematical or analytic truths can be justified even if the believer is unable to provide any supporting reasons. Moreover, it is not clear to me that simply providing the reasons that support a particular belief is sufficient for justifying that belief. Consider again the case of the identical gym bags. Even after providing all of the reasons that support my belief that “This bag is mine”—e.g., it’s got all of the features that I remember my bag having—it is still not obvious to me that my belief is justified in the presence of an equally likely, sufficient counter-weighing set of defeaters.

---

According to the second conception of demonstrative justification, a person is justified only if he is able to “adequately defend its status as justified.” This conception is more obviously problematic. Not only is it unclear what “adequately” will ever mean, but also, surely there are some who do not have the ability to adequately defend their own beliefs (e.g., children) and are yet still justified in holding at least some of those beliefs.

And finally, on the third conception of demonstrative justification, a person is justified in holding some belief only if he is able to persuade another of its truth, or of its justification. This, again, is far too high a standard for justification. If I am the only person to witness a particular event, surely my memorial beliefs about the event are justified whether or not I am able to convince others that it all happened as I remember it. I am justified in my belief that the earth orbits the sun even if I am unable to convince Gerardus Bouw that I am right, or even that I am justified.

So how else might a person be agent-justified if they are not demonstratively justified? Shafer-Landau's answer is to allow justification to be a matter of things external to the believing person. For instance, if, according to a fairly basic conception of reliabilism, the justification of a belief $B$ is a matter of its being the result of a reliable belief-forming process, then as long as $B$ was reliably formed, the person holding $B$ could still be justified even if she is unable to demonstrate why. And so perhaps the problem I’ve raised for moral knowledge on weak supervenience can be avoided by considering this distinction. Even if

---

78 Gerardus Bouw is the leader of the Association for Biblical Astronomy and editor of the ABA's newsletter, The Biblical Astronomer. He has a Ph.D. in Astronomy and is a devout geocentrist.
79 But wait, wasn’t he the one that just admitted the possibility of self-evident moral beliefs? Why the appeal now to externalism? According to Shafer-Landau, moral principles like “It is wrong to break promises” are self-evident, but verdictive moral beliefs like “It is wrong of Jones to break that promise” will require an externalist conception of justification. And so he calls his own moral epistemology and kind of “hybrid theory,” according to which “We keep the picture of verdictive belief that we have been developing [reliabilism], and supplement it with the claim that the process of adequately understanding certain propositions (the self-evident ones) is itself sufficiently reliable to confer warrant” (2003: 279).
we are not able to demonstratively justify our moral beliefs by appropriately distinguishing the actual world from \( W^* \), those beliefs may still be justified if they are the result of a reliable moral belief-forming process. In this sense, some naturalistic version of moral reliabilism might be effective as an answer to the problem I’ve been raising for moral justification given weak supervenience.

The problem for the ethical naturalist will be to explain how we get from the suggestion that our moral beliefs might have been reliably formed to an explanation of how they are reliably formed. After defending the plausibility of moral reliabilism, Shafer-Landau concedes, “Even if most of what I have said thus far is on the mark, we are still far short of knowing how to acquire moral knowledge.” Until the ethical naturalist either (a) defines which process of moral belief formation is reliable, or perhaps (b) offers a clear set of criteria for evaluating the reliability of such processes, all he really has is the possibility of moral knowledge, not moral knowledge. And I think that both (a) and (b) will turn out to be especially difficult. Shafer-Landau’s account of moral reliabilism strikes me as successful only to the extent that it relies first upon self-evident moral principles. And as I’ve explained, this is not an option available to any thoroughgoing naturalist. Furthermore, whatever account is given will not only have to deal with the problems I’ve raised for justification given weak supervenience, but also with the problems raised earlier by Street. There is good reason for the ethical naturalist to believe that evolutionary pressures have played a significant role in the formation of our moral beliefs. But if we’ve got good reason to believe that these evolutionary pressures were never truth-aimed to begin with, then we’ve got good reason to believe that our processes of moral belief formation have been largely unreliable.

---

80 Shafer-Landau 2003: 293.
3.4 Conclusion

I think this all leaves the ethical naturalist in a very uncomfortable position. On the one hand, as I argued in Chapter 1, strong moral supervenience pushes the thoroughgoing ethical naturalist in one of two directions: (a) he must either build unique facts and properties about divergent social structures (or varying moral belief systems) into his subvenient sets, and so jeopardize the objectivity of moral truths, or (b) he must insist, in the face of all possible worlds in which people have different moral beliefs than ours, that they are all mistaken—this despite the fact that the belief-forming mechanism responsible for his moral beliefs was never concerned with the truth of those beliefs in the first place. And on the other hand, as I’ve argued in this chapter, weak moral supervenience is a defeater for moral knowledge—that is, until the ethical naturalist offers a fully satisfying externalist account of moral justification.

My aim in this thesis has been to put the ethical naturalist “on the ropes,” as it were, by setting forth a prima facie case for the inadequacy of ethical naturalism. In this regard, I see my project as dovetailing rather nicely with the recent revival of non-naturalism in ethics. Whether or not there is a way for naturalists to respond to my worries about moral supervenience while still maintaining both (a) a thoroughgoing naturalism and (b) all of the essentials of moral realism, I will leave it to them to determine. My initial suspicion, however, is that much of the problem just is their thoroughgoing naturalism, and so some variety of either (a) ethical non-naturalism, or (b) anti-realism about moral properties\(^81\) may actually be their only options.

---

\(^{81}\) Indeed, Street’s conclusion in her paper is that Darwinian considerations and their implications for ethical naturalists offer us excellent reason to be anti-realists.
REFERENCES


