In memory of Fr. Thomas H. Green, S.J.,
who not only introduced me to both Wodehouse and Wittgenstein,
but also taught me what it means to be good.
Acknowledgments

This work would not have been completed without the help of my supervisors, Dr. Michael Pelczar and Dr. Kyle Swan. Dr. Pelczar was already working on a paper on the similarity between the open question argument and the knowledge argument when I came to NUS in 2005. When I showed him “A Lecture on Ethics” and said I want to work on the relevance of Wittgenstein’s ideas to ethics beginning with this lecture, he immediately pointed out that there is an interesting argument against moral naturalism in Wittgenstein’s description of a person who knows all natural facts but does not include any ethical statements in a complete description of the world. I have been working on this argument against moral naturalism since then. And I have come to appreciate the similarity between Wittgenstein’s argument against moral naturalism and the knowledge argument through Dr. Pelczar’s work on this topic and in the philosophy of mind. I did not work as much on Wittgenstein as I had originally planned but I more than made up for this by reading more on metaethics. Dr. Swan provided invaluable help in making sense of the expansive and sometimes confusing literature on the subject. He also gave innumerable comments and suggestions to the many drafts that I wrote along the way. Without Dr. Swan’s comments and suggestions I would have undoubtedly committed a lot more blunders in writing about moral naturalism.

I also acknowledge the unflagging patience and support of my family, without whom I would not have been able to complete this work. Thanks to my mother and father, and also to May, Yayan, Bim, and Mica.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. MOORE’S ARGUMENT AGAINST MORAL NATURALISM</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Moore’s open question argument</td>
</tr>
<tr>
<td></td>
<td>Some standard objections and replies</td>
</tr>
<tr>
<td></td>
<td>Towards a new argument against moral naturalism</td>
</tr>
<tr>
<td>3. INTRODUCING JEEVES: SOMEONE WHO KNOWS ALL NATURAL FACTS BUT DOES NOT KNOW WHETHER SOMETHING IS MORALLY GOOD</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Wittgenstein’s argument against moral naturalism</td>
</tr>
<tr>
<td></td>
<td>The world book argument and the open question argument</td>
</tr>
<tr>
<td></td>
<td>The world book argument and the knowledge argument</td>
</tr>
<tr>
<td>4. TAKING JEEVES SERIOUSLY: WOULD SOMEONE WHO KNOWS ALL NATURAL FACTS REALLY FAIL TO KNOW WHETHER SOMETHING IS GOOD?</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>The notion of fallible omniscience</td>
</tr>
<tr>
<td></td>
<td>A naturalist’s account of omniscience</td>
</tr>
<tr>
<td></td>
<td>Moral disagreement and knowledge of all natural facts</td>
</tr>
<tr>
<td>5. SIZING UP THE WORLD BOOK ARGUMENT: DOES FALLIBLE OMNISCIENCE REALLY IMPLY THAT MORAL NATURALISM IS FALSE?</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Natural properties versus moral properties</td>
</tr>
<tr>
<td></td>
<td>An ability hypothesis about Jeeves</td>
</tr>
<tr>
<td></td>
<td>Reply to the ability hypothesis about Jeeves</td>
</tr>
<tr>
<td>6. SECOND THOUGHTS ABOUT THE WORLD BOOK ARGUMENT: IS IT POSSIBLE TO ACCOMMODATE JEEVES WITHIN A SOPHISTICATED MORAL NATURALIST FRAMEWORK?</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>The new guise hypothesis about Mary</td>
</tr>
<tr>
<td></td>
<td>A new guise hypothesis about Jeeves</td>
</tr>
<tr>
<td></td>
<td>Reply to the new guise hypothesis about Jeeves</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>124</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>129</td>
</tr>
</tbody>
</table>
Summary

In “A Lecture on Ethics,” Ludwig Wittgenstein explicitly adopts some of G.E. Moore’s ideas: “My subject, as you know is Ethics, and I will adopt the explanation of that term which Professor Moore has given in his book *Principia Ethica*.” Despite this reference, Wittgenstein’s and Moore’s parallel views on ethics have not been explored or appreciated much. What is more, nobody saw that Wittgenstein also presents an improved version of Moore’s argument against moral naturalism. This dissertation seeks not only to demonstrate that Wittgenstein’s argument against moral naturalism is similar to Moore’s well-known open question argument; it also argues that Wittgenstein’s argument is better. This argument is based on the idea that someone who knows all natural facts would not know whether something is morally good. Wittgenstein’s argument anticipates and even improves on one contemporary version of Moore’s argument, which poses a problem for at least one type of moral naturalism. Despite some marked differences in the views of moral naturalists, most of them are inclined to say that someone who knows all natural facts would also know everything there is to know about moral goodness. This makes Wittgenstein’s argument against moral naturalism especially relevant for contemporary moral naturalists. Wittgenstein’s argument is also relevant to full information accounts of moral goodness, which are widely prevalent today. The imaginative possibility that Wittgenstein presents can be philosophically substantiated in a way that casts doubt on the adequacy of such full information accounts of moral goodness for vindicating moral naturalism.

In some respects Wittgenstein’s argument against moral naturalism is similar to Frank Jackson’s knowledge argument against physicalism.
Wittgenstein’s argument features someone who knows all natural facts but does not know whether something is morally good while Jackson’s argument features someone who knows all physical facts about color vision but does not know what it is like to see a red. The similarity between the two arguments and the fully informed characters in them will be explored through a discussion of a distinct type of moral disagreement. A precursor of this type of disagreement is featured in Terence Horgan and Mark Timmon’s revival of Moore’s argument against moral naturalism. In this type of moral disagreement the disputants agree about the natural features of the matter under discussion, but disagree about what moral conclusions to draw from these natural features. Because of this type of moral disagreement, someone who knows all natural facts can fail to know whether something is morally good. However, there is at least one important difference between Wittgenstein’s argument and Jackson’s. The latter is built around phenomenal properties, the existence of which cannot be doubted. The former, on the other hand is built around moral properties, which are by no means indubitable. Because of this difference, the knowledge argument and Wittgenstein’s argument against moral naturalism establish contrasting metaphysical conclusions. Whereas the knowledge argument establishes the discrete metaphysical existence of phenomenal properties, Wittgenstein’s argument establishes that moral properties are not natural in a way that casts doubt on the existence of moral properties.
Most people think that knowing whether something, say capital punishment or physician assisted suicide, is morally good depends on knowing a lot of facts about the moral matter in question. Roughly, the idea is that the more we know about the deterrent effect of capital punishment or all the different consequences that arise from legally permitting physician assisted suicide, among other things, the better equipped we shall be to judge whether it would be good to adopt policies that permit these. This general idea fits well with full information or ideal advisor accounts of moral goodness. According to those who propose these accounts, moral goodness is determined by what an individual, or a community composed of individuals, who are fully and vividly informed about themselves and their environment would want. Because of what they know these individuals are in a position to advise their non-idealized selves what to do. Sometimes the same idea is expressed in terms of what one would do if one were well-informed, calm, cool and collected.

But most people also think that being fully informed alone does not guarantee knowledge about moral goodness. If I say that it is possible for someone to be fully informed about the relevant moral matter, calm, cool and collected but ignorant about whether say capital punishment is morally good, nobody would balk at my imaginative proposal. Even philosophers who are partial towards full information accounts of moral goodness might be willing to accept this under-described proposal without many qualms. It is easy enough to come up with a story that could explain away the possibility of being fully informed about something yet ignorant about that thing in some morally relevant way. Or so it seems. It is both interesting and puzzling that there is a
tendency to think such an explaining away could be done easily without looking more closely at the merits of the different explanations that can be given. On closer inspection the possibility of knowing all natural facts but still being morally ignorant may not be so easy to dismiss. It could spell trouble not only for full information accounts of moral goodness but also for different forms of naturalistic moral realism, moral naturalism for short.¹ In this dissertation a new argument against moral naturalism built around the imaginative possibility just described will be discussed and defended.

The new argument against moral naturalism comes from a relatively old source: Ludwig Wittgenstein’s “A Lecture on Ethics.” Neither at the time of its delivery in 1929 nor in the years following its appearance in print in 1965 has this article affected debates on ethical theory. At the time when Wittgenstein’s ideas were fashionable the lecture on ethics together with Wittgenstein’s other unpublished writings were widely read; but only as one of Wittgenstein’s writings. After the height of interest in Wittgenstein’s so-called later philosophy, or what was then called ordinary language philosophy, the lecture on ethics became a forgotten period piece. It is possible that the argument against moral naturalism contained in the lecture escaped notice because the interest in Wittgenstein’s philosophy was already waning in the 1960’s,² when it was published in the Philosophical Review. What happened was unfortunate. Wittgenstein’s lecture contains some ideas that have much to contribute to some debates on ethical theory. This dissertation seeks not only to

---

¹ Since virtually all ethical theorists adhere to a naturalistic world view, the convention of using the term ‘moral naturalism’ to refer to those moral theories that hold that there are objectively existing moral facts and properties and that these are natural facts and properties will be adopted here. More on moral naturalism in the following pages.

demonstrate that the argument against moral naturalism from Wittgenstein’s lecture is similar to G.E. Moore’s well-known open question argument; it also argues that Wittgenstein’s argument is better.

The new argument against moral naturalism is Wittgenstein’s in the sense that it originates from some remarks that he makes in his lecture on ethics. It is unclear whether Wittgenstein himself intended his remarks about someone who knows all natural facts but does not know whether something is morally good to be taken as an argument against moral naturalism like Moore’s open question argument. It is also unclear how or whether the ideas presented in “A Lecture on Ethics” are related to statements about ethics in the Tractatus or to those remarks that have implications for ethics in the Philosophical Investigations. Since this lecture was given at the beginning of Wittgenstein’s so-called middle period—the time when he had already begun to turn away from ideas contained in the Tractatus but has yet to fully realize and develop the ideas that culminated in the Investigations—asking how it is related to his more well-known but diametrically opposed ideas is very thought-provoking. Inquiring further into these questions could prove to be very interesting and highly significant for Wittgenstein scholarship. The present work, however, does not pursue this inquiry. The main concern here is

3 There is no direct evidence in Wittgenstein’s published writings that indicates that he took his remarks in LE as a new version of Moore’s argument against moral naturalism.

4 What is sometimes referred to as Wittgenstein’s ‘middle period’ begins in 1929 when he returned to philosophy after having been a schoolteacher for some years and ends around the time he wrote what is known today as The Blue and Brown Books, which he finished dictating in 1935. This part of Wittgenstein’s philosophy is set apart from his late and early writings because “it gradually became clear that his writing during the intervening years ... could not simply be understood as a rejection of one view and the adoption of another ... [so it became] natural to speak of this further body of writing ... as ‘Middle Wittgenstein’.” [David Stern, “The ‘Middle Wittgenstein: From Logical Atomism to Practical Holism,” Synthese 87 (1991): p. 203.]
not exegetical. The discussion of Wittgenstein’s argument against moral naturalism that follows is mainly focused on philosophical considerations. Some ideas from “A Lecture on Ethics” are taken as a springboard for discussing a new and interesting way to argue against moral naturalism.

No discussion of moral naturalism, or indeed of metaethics is complete without any consideration of Moore’s open question argument. Thomas Baldwin goes so far as to say that twentieth-century British ethical theory is unintelligible without reference to Moore’s *Principia Ethica*. Although virtually everyone in the last eighty or so years rejects Moore’s moral non-naturalism, his argument against moral naturalism remains of interest to ethical theorists. After the heyday of ethical intuitionism during the first three decades of the last century, the open question argument was taken to show that moral naturalism is problematic because it cannot account for the prescriptive or emotive

---


6 G.E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903). Hereafter this work will be referred to as *PE*.

According to Thomas Hurka, *PE* came in “the middle of a sequence of ethical writing that runs roughly from the first edition of Sidgwick’s *Methods of Ethics* in 1874 to Ross’s *Foundations of Ethics* in 1939.” [“Moore in the Middle,” *Ethics* 113 (2003), p. 600.] From 1874 to 1939 the moral non-naturalism developed by Moore, Rashdall, Ross, et. al., was a widely accepted ethical theory. In the next twenty years it was rejected in favor of antirealist accounts of moral values, which were then called moral non-cognitivism (Baldwin, Ibid.). When philosophers began posing objections against non-cognitivism in the 1950s moral naturalism again gained many proponents. At first only Aristotelian forms of naturalism gained new ground; G.E.M. Anscombe, P.T. Geach and Philippa Foot all promoted this view. But later on philosophers like Richard Brandt, Richard Boyd, and many others developed different forms of moral naturalism. Today there are those who argue for a new kind of moral non-naturalism. But those who develop this view eschew many features of Moore’s metaethical position. See Russ Shafer-Landau, *Moral Realism: A Defence* (Oxford: Clarendon, 2003).
dimension of morality. At the same time it became evident that Moore’s metaethical position demands an extravagant metaphysics and epistemology.\(^7\) The existence of non-natural moral properties does not sit well with the world picture afforded us by the natural sciences. So until about 1960 considerations arising from the open question argument were accepted as reasons for going for an antirealist construal of moral facts and properties.\(^8\) Later developments in philosophy of language and in the notion of conceptual analysis, however, induced many to rethink the implications of the open question argument. In the last few decades some ethical theorists began to think that Moore’s argument against moral naturalism can be debunked in ways that show that moral naturalism is plausible.

At the very least, some contemporary philosophers think that what’s important about the open question argument has something to do with an insight about a unique feature of moral terms and concepts that is neutral between moral naturalism, which is a form of realism, and antirealism. In their landmark article entitled “Toward Fin De Siècle Ethics: Some Trends,” Dar-


\(^8\) J. L. Mackie took the most singular approach towards the extravagant metaphysical and epistemological demands of Moore’s moral non-naturalism. Mackie conceded the point about the peculiar evaluative, prescriptive, intrinsically action-guiding aspectes of non-natural moral properties. But he pointed out that this picture does not fit at all with our scientific, i.e. naturalistic, understanding of the world. Thus, according to Mackie, a position like Moore’s implies that there are no objective values. *Ethics: Inventing Right and Wrong* (London: Penguin, 1977), pp. 32-5.

wall, Gibbard and Railton refer to this feature of moral terms and concepts as normativity, or action-guidingness. They also say that the significance of the normativity of moral terms and concepts is really what Moore's open question argument uncovered. When we ask of any naturalistic analysis of moral goodness on offer whether it is good, we are really asking if, all things considered, we should devote ourselves to promoting whatever it is that such an analysis takes to be good. Darwall, Gibbard and Railton also say that the significance of Moore's argument against moral naturalism lies in its discovery of an argumentative device that "implicitly but effectively brings to the fore certain characteristics of 'good'." Since there is nothing that prevents naturalists from appropriating such characteristics for their own ends, Moore's argument does not really threaten moral naturalism.

Moral naturalists believe that moral properties are natural properties. They also believe that the correctness of sincerely uttering a statement that features a moral predicate depends on the instantiation of natural properties. As naturalists see it, true moral beliefs are acquired through empirical means. Moral naturalism combines the attractiveness of believing that there are moral properties with the advantages of adhering to a naturalistic world view. Belief in moral properties is vouched for by philosophical tradition and common sense, while adhering to a naturalistic world view is parsimonious. But natu-

10 "Toward Fin de siècle Ethics," p. 117.

11 Ibid.

12 After this point I will only talk about moral properties. If moral naturalism is true, true moral facts and properties are natural facts and properties. A good candidate for a moral fact is 'Slavery is wrong.' This statement attributes the property of moral wrongness to slavery. So a moral fact expresses an instantiation of a moral property. Moral properties denote an objective quality or attribute of persons, actions or states of affairs. So talk of moral properties is shorthand for all these, and of moral facts.
Some naturalists think that moral properties are reducible to natural properties; but others deny this. Richard Brandt and Peter Railton are among those who think that moral properties are reducible to natural ones. According to them moral properties are coextensive with natural properties. Railton points out that it would be odd if moral properties were natural but also *sui generis*.\(^{13}\) He goes on to say that what is morally good is determined by facts about “what would rationally be approved of were the interests of all potentially affected individuals counted equally under circumstances of full and vivid information.”\(^{14}\) In the same vein, Brandt says that what is morally good can be identified with “which state of affairs would be wanted, in a long-range view, for their own sake by a person who was fully factually informed.”\(^{15}\) Although he has different views on conceptual analysis, Frank Jackson also believes that moral properties are reducible to natural ones. Unlike Brandt and Railton, Jackson thinks that moral properties are a function of what mature folk morality takes to be relevant in determining which morally evaluative properties are which descriptive properties.\(^{16}\)

On the other hand, Richard Boyd, David Brink and Nicholas Sturgeon believe that although moral properties are natural properties, the former are not reducible to the latter. The non-reductive moral naturalism of Boyd, Brink and Sturgeon consists in the belief that there are facts ‘as hard as any in


arithmetic or chemistry’ that there are certain rules of conduct and dispositions of character that are appropriate for human beings.\textsuperscript{17} Despite these differences in the moral naturalist camp, most naturalists are inclined to say that someone who knows all natural facts would also know everything there is to know about moral goodness. Most moral realists would likewise be inclined to say that someone omniscient about natural facts would be omniscient about moral matters as well. This confluence of opinion on what someone who knows all natural facts would know about moral matters makes an argument based on fallible omniscience effective against virtually all forms of moral naturalism and most versions of moral realism.

Antinaturalists deny that moral properties are natural. The argument against moral naturalism from Wittgenstein’s “A Lecture on Ethics” takes issue with the claim that someone who knows all natural facts would always know whether something is morally good. Wittgenstein thinks that it is possible for a person to know all natural facts without knowing whether something is morally good. For the sake of brevity, call this possibility fallible omniscience and the compilation of all natural facts a world book. Darwall, Gibbard and Railton identify Wittgenstein’s lecture on ethics as the locus classicus of the view of normativity or action-guidingness that they discuss in connection with the open question argument. They note that unlike Moore, “Wittgenstein ... could see that ‘no description that I can think of would do to describe what I mean by absolute value, [and] I would reject every significant description anybody could suggest.’”\textsuperscript{18} They also think that both Wittgenstein’s and Moore’s arguments against moral naturalism depend on the normativity or the


\textsuperscript{18} “Toward Fin de siècle Ethics,” p. 119.
action-guidingness of moral terms and concepts. According to them, both arguments trade on an insurmountable difference between saying that something has certain natural properties and saying that this thing is morally good. This insurmountable difference depends on a semantic platitude: whereas moral predicates have action-guidingness or normativity logically built into them natural ones do not. Darwall, Gibbard and Railton’s assessment of both Moore’s and Wittgenstein’s arguments against moral naturalism may have something to do with the continuing neglect that the argument against moral naturalism based on fallible omniscience has suffered. But fallible omniscience is not necessarily tied up with the normativity of moral terms and concepts. The similarity between someone who knows all natural facts but does not know whether something is morally wrong with someone who knows all physical facts about color vision but does not know what it is like to see red implies this.

In some respects Wittgenstein’s argument against moral naturalism is similar to Jackson’s knowledge argument against physicalism. The similarity between these two arguments and the fully informed characters in them will be explored to the antinaturalist’s advantage here through a discussion of a distinct type of moral disagreement. A precursor of this type of disagreement is featured in Terence Horgan and Mark Timmon’s revival of Moore’s argument against moral naturalism. The discussion of Moore’s argument in the next chapter will be focused on how an argument similar to Moore’s can work better if it relies on this type of moral disagreement rather than on questions of the sort that Moore posed.

Moral naturalism is the default position among moral realists, those who believe that (a) there are moral properties; (b) some moral beliefs are true; and (c) true moral beliefs are made true by the instantiation of moral
properties. Some moral realists support their position by drawing an analogy between philosophy of mind and metaethics. Roughly, the idea is to show that moral properties resemble conscious experience in important ways. Non-reductive physicalism about phenomenal properties associated with conscious experience combine many of the features that a moral realist needs to defend her claims about moral properties. Jackson, McDowell and Shafer-Landau make use of an analogy between philosophy of mind and metaethics in different ways.\(^\text{19}\) The metaethical position of these philosophers are by no means similar but they all make use of a similar strategy. Jackson appeals to the supervenience on the descriptive and makes use of strategies used in the philosophy of mind to establish the supervenience of the mental on the physical to show that moral supervenience implies that moral properties are natural properties.\(^\text{20}\) McDowell draws an analogy between perceiving moral properties and perceiving color to establish his own brand of moral realism.\(^\text{21}\) Shafer-Landau makes use of the notion of intensional individuation to establish a non-naturalist moral metaphysics that nevertheless concedes that the moral supervenes on the descriptive.\(^\text{22}\)

But the strategy of drawing an analogy between philosophy of mind and metaethics is neither simple nor straightforward. There is at least one important dis-analogy between phenomenal properties and moral properties:

\(^{19}\) William G. Lycan also talks about the possibility of finding a way to argue for an \textit{a posteriori} reduction of moral properties into natural properties by appealing to some of the strategies that philosophers use in their work on the mind-body problem.


\(^{20}\) \textit{From Metaphysics to Ethics}, pp. 118-25.


\(^{22}\) \textit{Moral Realism}, pp. 80-115.
the existence of the former is indubitable while the existence of the latter is not. Because of this difference the knowledge argument and Wittgenstein’s argument against moral naturalism establish contrasting metaphysical conclusions. Whereas the knowledge argument establishes the discrete metaphysical existence of phenomenal properties, Wittgenstein’s argument establishes that moral properties are not natural in a way that casts doubt on the existence of moral properties. The significant difference between phenomenal properties and moral properties will be used to defend the crucial premise of Wittgenstein’s argument, which states that fallible omniscience implies that moral properties are not natural. In this way the new argument against moral naturalism would also prove to be effective against the different kinds of moral realism mentioned above.

The structure of this dissertation is straightforward: Chapter 2 is devoted to a discussion of Moore’s argument against moral naturalism. The new argument against moral naturalism is introduced in Chapter 3. The first premise of this argument is discussed in Chapter 4 while the exposition and defense of the second premise are taken up in Chapters 5 and 6.
Chapter 2
Moore’s Argument Against Moral Naturalism

One of the most well-known arguments against moral naturalism is Moore's open question argument (OQA). In this chapter an exposition of Moore's argument will be given in order to show that an argument similar to it could prove to be a formidable challenge to some contemporary forms of moral naturalism. The open question argument is as well-known as it is problematic. Numerous objections have been raised against this argument and it is generally agreed that questions of the sort Moore posed to make a case against moral naturalism do not provide sufficient grounds for a decisive antinaturalist conclusion.¹ But such a conclusion may still be forthcoming if an argument similar to Moore's focuses on a distinct type of moral disagreement and what such a disagreement implies for the moral knowledge of a person who knows all natural facts. The discussion of this new argument

¹ Even Moore later admitted that all his supposed proofs that moral goodness is not natural 'were certainly fallacious.' [Philosophical Papers (London: Allen and Unwin, 1959), pp. 89-101.] He also admitted that his attempt to explain what a natural property is in the Principia Ethica was 'hopelessly confused.' ['Reply to My Critics,' The Philosophy of G.E. Moore (La Salle: Open Court, 1968), p. 582.]

Today, philosophers who subscribe to diverse ethical theories say more or less the same thing about Moore's argument against moral naturalism. David O. Brink argues that Moore relies on an implausible 'semantic test for properties,' according to which two terms that are not analytically equivalent cannot refer to the same property. Brink points out that this is clearly a false assumption because we make discoveries that flout this assumption all the time, e.g. Water = H₂O, genes = DNA, etc. Moral Realism and the Foundations of Ethics (Cambridge: Cambridge University Press, 1989), pp. 159-67. Gilbert Harman argues that the open question argument is inconclusive for similar reasons; an argument analogous to Moore's can be used to prove that water is not H₂O. The Nature of Morality: an Introduction to Ethics (New York: Oxford University Press, 1977), pp. 17-20. Michael Smith points out that the force of the open question argument depends on the dubious assumption that analyses cannot be correct if they are also unobvious and informative. The Moral Problem (Oxford: Blackwell, 1994), pp. 35-9.
against moral naturalism will be taken up in the next chapter.

Moore’s Open Question Argument

There have been innumerable discussions of Moore’s argument. Although various commentators offer different versions of the open question argument, the best account of the core of the argument is still the one in *Principia Ethica*:

If, for example, whatever is called ‘good’ seems to be pleasant, the proposition ‘Pleasure is the good’ does not assert a connection between two different notions, but involves only one, that of pleasure, which is easily recognised as a distinct entity. But whoever will attentively consider with himself what is actually before his mind when he asks the question ‘Is pleasure (or whatever it may be) after all good?’ can easily satisfy himself that he is not merely wondering whether pleasure is pleasant. And if he will try this experiment with each suggested definition in succession, he may become expert enough to recognise that in every case he has before his mind a unique object, a distinct question may be asked. Everyone does in fact understand the question ‘Is this good?’ When he thinks of it, his state of mind is different from what it would be, were he asked ‘Is this pleasant, or desired, or approved?’

Moore puts an emphasis on the mental state of competent speakers when they ask questions like ‘Is pleasure good?’ because what propels the argument is the distinction between the concept of moral goodness and the concept of having some natural property. Moore thinks that this conceptual distinction is sufficient to establish the discrete metaphysical existence of moral properties and natural properties. Since open questions demonstrate this distinction he thinks that it is a powerful polemic against the naturalist.

---

2 *PE* 16-7

3 The concept of moral goodness is associated with prescriptions for action or imperatives whereas the concept of having some natural property is not. The latter has more to do with the descriptive rather than the prescriptive. It is unclear whether Moore meant to highlight this distinction when he spoke of the difference between the mental state of competent speakers when they ask ‘Is pleasure good?’ and the mental state of the such speakers when they ask ‘Is pleasure pleasure?’
The argument can be stated thus:

OQ1  For any natural property, it is an open question whether something which has this natural property is morally good.

OQ2  If for any natural property, it is an open question whether something which has this natural property is morally good, then moral naturalism must be false.

OQ3  So moral naturalism is false.

An open question is one that a competent speaker can ask without betraying any conceptual confusion. The best way to show what this definition amounts to is to give an example of a closed question: is a brother a male sibling? This question is closed because the answer to it should be obvious to anyone who understands the terms 'brother' and 'male sibling.' The concepts involved in this question are synonymous; anyone who grasps them and is sufficiently familiar with the language in which the question is posed knows that the only conceivably correct answer is “yes.” No further investigation is necessary in order to answer a closed question. Stating the meaning of the terms involved in the question is enough to show anyone with sufficient understanding that the answer is right before her eyes. In contrast, there are no quick and easy answers to open questions. Answering them may require further investigation, even if we already grasp all the concepts required to understand the question. Not knowing the answer to an open question betrays an ordinary lack of information, as opposed to linguistic incompetence or conceptual confusion.

Moore defends the first premise of his argument by showing that naturalistic definitions of moral goodness are prone to open questions. Moore claims that “whatever definition be offered, it may be always asked, with
significance, of the complex so defined, whether it is itself good."4 For instance, since hedonists claim that goodness is nothing other than pleasure, we can ask, 'But is pleasure good?' This question is open because it is patently different from 'But is pleasure pleasure?' Knowing the meaning of both 'good' and 'pleasure' is not sufficient to answer the question 'But is pleasure good?' Even someone who is inclined to agree with the hedonist might say that further explanation and argument are needed to show that pleasure is indeed good. Defining moral goodness as 'what we desire to desire' is open to the same objection. 'But is what we desire to desire good?' is clearly not the same thing as 'But is what we desire to desire what we desire to desire?' The first question is open, the second is not.

Moore thinks that questions of the sort he raises against naturalistic definitions of 'good' are open because each involves 'two distinct notions.' He takes the openness of such questions to be a sufficient basis for the conclusion that having the property of moral goodness is distinct from having some purely descriptive natural property.5 Moore's explanation can be taken to show that the first premise of his argument applies to a wider range of natural properties, even complex or sophisticatedly constructed ones. If someone proposes a complex naturalistic analysis of 'good' we can ask of the focal point of such an analysis whether it is good. For instance, if someone says that 'good' can be defined in terms of a sociobiologically determined altruism—character traits or behavioral dispositions that developed through natural selection which promote the flourishing of our species—we can still

4 PE p. 15

5 This of course is one of the most contested moves in Moore's argument, especially in light of the more sophisticated forms of naturalism endorsed by philosophers today. Open questions might demolish analytic naturalism but other forms of naturalism can accommodate questions of the sort Moore raises.
A plausible story can be told about how 'good' and other ethical concepts are a byproduct of evolution. Traits and dispositions like being predisposed to make friends and practices like distributing goods on the basis of both need and merit helped humans to survive, so the disposition to abide by similar traits and practices were called 'good.' When we ask questions like 'But is such a sociobiologically determined altruism good?' and 'Are the traits we acquired through natural selection good?' it becomes apparent that these questions are not closed. They make sense in a way that 'Is a brother a male sibling?' does not. When someone asks these questions she is not asking if the traits and dispositions in question do in fact make us fit for survival; she already understands that they do. Perhaps she is asking if we, as a biological species, should evolve further. Whatever she may have in mind, the important point is that a naturalistic description of things that we take to be morally good still does not capture something important about morality. So much so that an analytic equivalence between goodness and a complex natural property appears inadequate. These considerations might seem like a good reason to eschew any position that posits analytic equivalence between moral properties and natural ones. In any case, Moore’s argument is directed against any kind of naturalism that depends on analytic identity between moral properties and natural ones.

Support for the second premise of the argument comes from Moore’s discussion of what he calls the naturalistic fallacy. This fallacy consists in conflating having the property of moral goodness with having some natural property, X. Moore complains that “far too many philosophers have thought

---

6 Asking this question points to the normativity or action-guidingness of moral terms like ‘good.’ But since Moore himself did not explicitly gloss open questions in this way normativity will not be discussed at length here.
that when they named ... other properties they were actually defining good." Moore concludes that it is wrong to identify goodness with any natural property, even if there is an objective correlation between some complex natural property and goodness. Moore endorses moral non-naturalism, but what he says about the implausibility of positing analytic identity between natural properties and moral ones fits well with saying that moral properties are natural but irreducible to non-moral or descriptive properties. But unlike the non-reductive naturalist Moore believes that when we ascribe goodness to something we are ascribing one simple property to it. As far as Moore is concerned, goodness is sui generis. In any case, Moore's non-naturalism resembles non-reductive naturalism because it holds that it is not possible to replace an attribution of goodness with an attribution of any natural property.

To say that something is good is different from saying that it is X, even if X were a complete description of something good in terms of its natural properties. This implies that there is something going on in attributions of moral goodness that is missing from attributions of any purely natural property. Even a complete natural description of something Moore takes to be morally good would not (he and the non-reductive naturalist think) capture what is going on when he attributes moral goodness to that something. Whether a naturalist can accommodate this discrepancy between attributions of natural properties and attributions of moral goodness is a separate question. Moore seems to think that only a negative answer to this question is plausible. But a separate argument is needed for this position.

---

7 PE p. 10
**Some standard objections and replies**

There are a number of well-known objections to Moore's argument against moral naturalism. I will discuss some of these objections and the ways in which they could be met or circumvented. The first premise of the argument is standardly charged with question begging. Those who pose this objection say that it depends on the significance of open questions in a way that gives the game away to the non-naturalist. They say that naturalists cannot grant that the natural property with which they propose to define goodness is open to the question 'But is it good?' in any philosophically significant way. For instance, a hedonist might be able to admit that she has to explain what she means when she says that goodness is nothing other than pleasure and why she thinks that this definition of moral goodness is right. But she also believes that after she has given a suitable account of pleasure and a justification of it, questions like 'But is pleasure good?' are already substantially answered, if not altogether silenced.

Thomas Baldwin offers a reply to the question begging objection on behalf of Moore. Baldwin points out that open questions do not have to be objectionable to the naturalist because they can claim that upon conceptual reflection competent speakers will stop asking question like 'But is pleasure...'

---

8 Those who pose this objection say that Moore supports OQ2 by implicitly assuming that moral goodness does not coincide with any natural property whatsoever. This is most evident in the discussion of the naturalistic fallacy. Indeed, this fallacy is the prime target of the charge Moore begs the question against his naturalist opponents. Frankena contends that by calling any attempt to define goodness in terms of natural properties a naturalistic fallacy, Moore assumes what he is trying to prove. The so-called fallacy that Moore coined can only be wielded if some other argument is given to knock down moral naturalism, something which Moore has not done. [W.K. Frankena, “The Naturalistic Fallacy,” *Mind* 48 (1939): 464-77]
good?" If the hedonist is right then upon conceptual reflection most, if not all, competent speakers would see that 'Is pleasure good?' is closed in the way that 'Is a brother a male sibling?' is closed. The proposed analysis of goodness "should come to seem to us entirely appropriate to guide our thought and judgments in accordance with it, even if at first the analysis strikes us as unobvious." For instance, after the hedonist gives an account of and an argument for her position most if not all of us ought to be satisfied already. But most of us aren't. And Moore claims that we do not find the naturalist's analysis of goodness acceptable, even upon conceptual reflection. The main insight behind Moore's open questions is that the more we think about it the more we see that 'good' and 'pleasure,' or any other proposed naturalistic definition, are two distinct concepts. So upon conceptual reflection we realize that an analytic moral naturalism cannot be true. The kind of questions that Moore raised shows that synonymy or identity between goodness and some natural property is impossible to establish. A different kind of naturalism might turn out to be true but the analytic kind is implausible. But a different form of naturalism will only fare better compared to the analytic kind if it can give a satisfying explanation for the insufficiency of natural analytic definitions for moral goodness.

Another standard objection to Moore's argument is that it presupposes an outdated semantics. This objection is related to the previous one in the following way: even if Baldwin's reply on behalf of Moore is true, the notion of analysis presupposed by his argument is mistaken so that the openness of

---


10 Ibid.
questions that he poses do not disprove the naturalist's position. Those who pose the outdated semantics objection go one step further, they say that if we rely on the natural term semantics of Frege, Kripke, and Putnam, we will see the real reason why questions of the sort Moore raises are open. There is usually more than one concept associated with a single referent. And most of the time we only become apprised of the fact that two concepts have the same referent through scientific discovery. Moreover, the more precise scientific concept to which one of our ordinary notions are equivalent is usually the result of conceptual and terminological innovation. For instance, before the discovery of what atoms in which specific ratio chemically composed water, it was thought that the colorless, transparent, odorless, tasteless liquid in seas, lakes, rivers, rain and most living organisms on earth is basic and indivisible—an atom in the literal sense of the word. Even after hydrogen and oxygen were discovered it was not obvious that the water molecule is composed of two hydrogen atoms and one oxygen atom (H₂O). Before 1826 'Is water H₂O?' was not only an open question, it was a live problem in scientific research. But when a more precise chemical nomenclature was proposed and instruments that could measure elements

---

11 The paradox of analysis objection is related to the objection from new moral semantics. According to C.H. Langford, Moore's open questions trade on a lexical accident which gives rise to the paradox that no conceptual analysis can be both correct and informative. It just happens that in English there is no synonym for 'good' in the same way that 'male sibling' is a synonym for 'brother.' But if an analysis of a concept is only correct in the way that 'brother = male sibling' then no new information can come from conceptual analysis. So as long as Moore does not allow for a more permissive kind of conceptual analysis questions like "Is pleasure good?" will persist. But if we accept the kind of analysis of concepts that natural science affords then there is no reason to think at the outset (as Moore seems to do) that no naturalistic analysis of moral concepts is plausible.

with greater accuracy were made it was discovered that the water molecule is composed of two atoms of hydrogen and one atom of oxygen. Now we take the correctness of ‘water = H₂O’ for granted; not because conceptual reflection allowed to see that it is right but because empirical discovery established it.

Those who propose to apply natural term semantics to moral terms think that ‘Is pleasure good?’ is open in the same sense that ‘Is water H₂O?’ was open to all competent speakers before 1826 (or is open to some competent speakers today who have yet to learn basic chemistry). They also say that unless we understand this new semantics properly we might be inclined to say that water is not H₂O, or that moral goodness is non-natural just because ‘water = H₂O’ is not an a priori truth. But this is clearly absurd, so Moore’s argument must be wrong-headed. And some kind of moral naturalism that eschew a priori equivalence between moral and natural properties could turn out to be right.

How can one reply to the objection from new moral semantics on behalf of Moore? A proponent of the open question argument must show that the proposed analogy between ethics and natural science is flawed. After all, this objection undermines the argument only if the semantics of natural kinds applies to moral predicates in the way that the naturalist needs to vindicate her theory. To show that there is an important difference between natural kinds and moral predicates it is necessary to look a little more closely at the former. ‘Water’ is a good example of a natural kind term. Hilary Putnam argues that the meaning of such terms “are not in the head” by presenting a twin earth thought experiment.¹² Twin earth is exactly like earth except that

the liquid they call water there is not H\textsubscript{2}O but a complex chemical compound that we could briefly summarize as XYZ. Just like on earth 'water' fills oceans, and lakes, falls to ground as rain, etc., in twin earth. When an earthling visits twin earth she might discover that on twin earth 'water' means XYZ. Likewise, when a twin earthling—a doppelgänger of the former—visits earth she could discover that on earth 'water' means H\textsubscript{2}O. But before such an interplanetary exchange, an earthling and a twin earthling would most likely disagree about the meaning of 'water.' One would say that water is H\textsubscript{2}O while the other would claim that water is XYZ. Once they closely examine the chemical composition of the liquid they both call water this disagreement would vanish. Putnam thinks that the twin earth thought experiment shows that natural kind terms like 'water' are rigid indexicals. 'Water' is a rigid designator in that its reference is fixed by our actual experience of H\textsubscript{2}O in our world. And it is an indexical in that the natural and socio-linguistic environment in which it is uttered determines its referent. If the new moral semanticists are right then moral predicates like 'good,' 'right,' etc., should behave just like 'water' in Putnam's twin earth scenario.

Non-reductive naturalists who endorse new moral semantics probably think that Putnam’s semantic externalism can help to establish the external or objective basis of our moral predicates. If the meaning of our moral terms are independent of competent speakers’ intrinsic states and our application of these terms co-vary with the natural features of the things to which we apply them, then they must refer to natural properties. These naturalists think that if they show that moral predicates behave like natural kind terms they could
make headway in proving that moral properties are objective and natural.\textsuperscript{13} The resulting form of naturalism would be immune from the open question argument because it can give a plausible naturalistic reason for the openness of the question ‘But is it good?’ Such questions are open because naturalistic equivalence or correspondence between two terms can only be established through empirical research and most of the time conceptual and terminological innovation is needed before we can even speak of such an equivalence, e.g. Water = \( \text{H}_2\text{O} \), temperature = mean molecular kinetic energy, etc. So there could be a natural property to which moral predicates refer even if we have not discovered it yet.

Terence Horgan and Mark Timmons offer a new version of Moore’s open question argument that could prove useful for replying to the objection posed by the new moral semanticists.\textsuperscript{14} They present a Putnam-style thought experiment that aims to show that moral terms do not rigidly refer to natural properties, even if these properties causally regulate our use of moral terms.\textsuperscript{15} Horgan and Timmons ask us to conceive of a planet exactly like ours in all natural respects. But whereas on earth moral terms are causally regulated by

\textsuperscript{13} If semantic internalism, the belief that what a speaker means by the words that she speaks depends on her intrinsic features, is true for both natural kinds and moral predicates, then moral naturalists could still use this semantic strategy to justify their position. But this time they have to forgo the objective or external component of their view. A moral naturalism that endorses semantic internalism about moral predicates would have something in common with moral subjectivism.


\textsuperscript{15} “Our central claim will be this: ...moral terms do not rigidly refer to the natural properties that causally regulate their use by humans. Although causal regulation may well coincide with--or even constitute--reference for certain terms (e.g. Names and physical natural-kind terms), we claim that for \textit{moral} terms anyway, causal regulation does not coincide with reference. I.e., moral terms do not refer to the natural properties that (we are supposing) causally regulate their use by humans (Ibid. p.160).”
consequentialist properties, on moral twin earth moral terms are causally regulated by nonconsequentialist properties. The notion of causal regulation of a moral term by some natural property or properties comes from Richard Boyd’s account of moral naturalism. According to Boyd, “a term t refers to a kind (property, relation, etc.) k just in case there exist causal mechanisms whose tendency is to bring it about over time, that what is predicated of the term t will be approximately true of k.” 16 This theory of reference allows us to say that premodern thinkers who were ignorant of the chemical composition of water, gold, etc., were nevertheless referring to these natural kinds that we pick out with greater precision today. Although Aristotle, for instance, would be unable to say that water and heavy water are the same compound by investigating physical properties of samples alone, his use of the term “water” (or more accurately, hudōr) is intended to be responsive to distinctions in nature that he may be mistaken about. Since ‘water’ is causally regulated by a specific chemical natural kind (H\textsubscript{2}O on earth) even at the time when what is true of H\textsubscript{2}O was unknown, the term ‘water’ already stands in the right relation with the natural kind that causally regulates its use. A naturalist like Boyd has it that moral terms could turn out to be like ‘water’ and other natural kind terms.

Horgan and Timmons’s moral twin earth scenario poses a challenge to Boyd’s position by showing that our semantic intuitions about a disagreement between earthlings and twin earthlings about some moral matter is different from our semantic intuitions about an analogous disagreement that features natural kinds. When a chemist and a twin chemist disagree about the chemical composition of water, the former saying that it is H\textsubscript{2}O while the latter in-

sists that it is XYZ, our semantic intuitions incline us to say that their disagreement is merely spurious. When they discover that the chemical composition of the liquid in rivers, oceans, clouds, etc., on earth is H\textsubscript{2}O but XYZ on twin earth their disagreement would cease to exist. If they are philosophically inclined they might event think that ‘water’ is causally regulated by some specific chemical kind at the planetary or global level, but more generally refers to a physical natural kind--a genus--of which H\textsubscript{2}O and XYZ are member species. But when an earthling and a twin earthling disagree about a moral matter our semantic intuitions do not incline us to say that their disagreement is merely spurious. Indeed, most of us are inclined to say that theirs is a genuine moral disagreement. Horgan and Timmons think that this marked difference between our semantic intuitions about moral terms and natural kind terms show that moral terms, unlike natural kind terms, are not referential and that synthetic moral naturalism is false.\textsuperscript{17}

Naturalists reply to Horgan and Timmons’s argument against moral naturalism by pointing out that our semantic intuitions about moral twin earth

\textsuperscript{17} Roughly, Horgan and Timmons’s argument against moral naturalism can be summed up as follows:

\begin{enumerate}
\item P1 If metaphysical moral naturalism is true then either it is analytic or it is synthetic
\item P2 If analytic moral naturalism is true, then questions of the sort that Moore raised must be closed.
\item P3 If synthetic moral naturalism is true, then the moral twin earth scenario must be interpreted (with the help of our semantic intuitions) in the same way that we interpret Putnam’s original twin earth thought experiments.
\item P4 But questions of the sort Moore raised are open.
\item P5 Our semantic intuitions incline us to interpret the moral twin earth scenario differently.
\item P6 So metaphysical moral naturalism must be false.
\end{enumerate}
are underwritten by the normativity of moral terms. They say that the reason a disagreement between earthlings and twin earthlings about some moral matter is genuine despite the marked difference between the natural properties that causally regulate their use of terms like ‘good’ and ‘right’ is that these terms are action-guiding.\(^{18}\) Regardless of which set of natural properties speakers associate with moral terms, they take these terms to give them reason to act in accordance with what they take to be ‘good’ or ‘right.’ So even if speakers do not use these terms to refer to the same set of natural properties, they could set this difference aside and have a meaningful argument about what ought to be done. The possibility of having such an argument hardly shows that moral terms do not function referentially; it merely shows that moral terms are not only referential in function. Since naturalists who subscribe to new moral semantics think that Moore’s argument trades on the normativity of moral terms and concepts so all that it really shows is that there is a difference between the descriptive and the normative.\(^{19}\) Since some of these naturalists account for moral normativity by endorsing motivational


\(^{19}\) This difference can be taken as good reason to eschew reductive analysis of moral terms and concepts. But it does not necessarily show that moral properties are not natural.
judgment externalism, rather than talking about moral semantics, their position is immune from both Moore's and Horgan and Timmons's antinaturalist arguments.

Towards a new argument against moral naturalism

If there is a type of moral disagreement that casts doubt on the usefulness of the notion of causal regulation for explaining away questions of the sort posed in the open question argument, then new moral semantics and a sophisticated moral psychology may not be enough to prevent an antinaturalist to set up a new argument similar to Moore's. The naturalist has it that like natural kind terms, moral terms refer to some unique set of natural properties even though we do not yet have a very clear idea of what these natural properties are. Even though we use terms like ‘good’ and ‘right’ in the same way that Aristotle used the term ‘water’--that is, crudely and without much empirical precision--at some point in time a full-fledged naturalistic moral theory would allow us to use moral terms with greater sophistication and accuracy. But if a certain type of moral disagreement stands in the way of discovering which set of natural properties moral terms really refer to, then the moral

20 Motivational judgment externalism is the belief that the connection between making a moral judgment and being motivated to act according to that judgment is contingent rather than necessary. Most externalists qualify this statement by saying that such a contingent connection could turn out to depend on human beings' deep or widely shared (natural, i.e. psychological) features. Naturalists who endorse externalism typically say that on their view being motivated to act according to a normative moral judgment can be widespread and predictable “even if it is neither necessary, nor universal, nor overriding (Brink 1989, p. 49).” Hence a moral naturalist can explain away the disagreement between earthlings and twin earthlings by saying that the difference between their usage of moral terms and concepts must be correlated with a difference in their psychological features. In the same way that in Putnam's twin earth the twin earthlings' bodies are made up mostly of XYZ rather than H₂O, people's psychological features on moral twin earth are constituted differently compared to earth's. But this hardly shows that moral naturalism is in trouble in the way that Horgan and Timmons say that it is.
naturalist’s hope of vindicating their position cannot be fulfilled, and a new antinaturalist argument might work.

The type of moral disagreement relevant here is similar to the one found in Horgan and Timmons’s response to the objection from new moral semantics. When an earthling and a twin earthling disagree about some specific moral matter, say capital punishment, their disagreement could persist even when they discover that their use of moral terms are causally regulated by different sets of descriptive properties, or even when they discover that different psychological features stand behind their use of moral terms. And as Horgan and Timmons pointed out, most people who contemplate such a situation would agree with the earthling’s and twin earthling’s assessment of their situation. Let’s assume that the earthling Oscar, whose usage of moral terms is causally regulated by consequentialist properties, thinks that capital punishment is morally justifiable under certain circumstances. Assume also that twin Oscar, whose use of moral terms is causally regulated by nonconsequentialist properties, thinks that there are no circumstances under which capital punishment is morally justifiable. When Oscar and twin Oscar argue about their conflicting moral beliefs about capital punishment it is difficult to

\[21\] In their article Horgan and Timmons stipulate that use of moral terms on twin earth are causally regulated by some nonconsequentialist property whose functional essence is captured by some specific deontological theory (p. 164). In the account of the disagreement between earthlings and twin earthlings given here I am merely relying on the coherence of saying that some specific form of deontological theory is compatible with the statement ‘There are no circumstances under which capital punishment is morally justifiable’ and that some specific form of consequentialism is compatible with the statement ‘There are some circumstances, namely those that maximize utility, in which capital punishment is morally justifiable.’ Of course it is possible to use consequentialist considerations to argue against capital punishment just as it is possible to argue for capital punishment on deontological grounds. What matters for the discussion here is that absent an assumption that some specific form of a normative theory, either consequentialist or nonconsequentialist, is correct, it is difficult, if not impossible, to say which of two moral theories that are internally consistent and at the same time congruent with descriptive facts is correct.
know which of them is right. Because they won't only be arguing about what ought to be done in this or that particular case that falls within the purview of their debate, but also about whose usage of moral terms is better, it is difficult to see how the debate can be settled. Like the hypothetical moral agent cited above who wants to know whether a sociobiologically determined altruism is good, the two Oscars want to know which, if any, of them is right. The intrac-
tability of their disagreement does not imply that moral naturalism is false, but such a disagreement could be used as a basis for saying that someone who knows all natural facts could be morally ignorant.

The two Oscars are liable to cite descriptive or natural facts which are true on both earth and twin earth to support their position. It is also likely that each would give a good argument for the internal consistency of their respective positions. Since the world which they inhabit have virtually the same descriptive features, their disagreement would not hinge on the correctness of factual reports but on the relation of these facts to the moral principle to which they subscribe and to their established use of moral terms. Oscar and twin Oscar might appeal to the internal consistency of their use of moral terms and the coherence of that usage with natural facts. But absent some way of arbitrating between two equally coherent naturalistic positions it is difficult, if not impossible, to settle the disagreement between them. Someone fully informed about the debate between the two Oscars might think that the moral norms to which the disputants subscribe and the semantic norms that govern their use of moral terms are irreconcilable. So much so that even access to more descriptive facts may not be enough to determine which moral norms and which semantic norms about moral terms, if any, are justified.

It is noteworthy that the hypothetical disagreement described above is already implicit in the contemporary debate on capital punishment. Pro-
tracted and impassioned discussion of the pros and cons of capital punishment has not yielded any clear indication of which side is right.\textsuperscript{22} Neither has it become any clearer how the dispute can be settled. It seems that nothing short of a sudden change in the evaluative priorities\textsuperscript{23} of one side could generate a consensus between the parties involved. It also seems that the best way to understand the persistent disagreement about capital punishment is to treat it not only as a dispute about what ought to be done but also as a difference of opinion about what counts as good. In other words, those who disagree about capital punishment may not only accept different morally relevant rules or principles for action, they may also be using ‘good‘ in different ways as a consequence of the difference between the moral norms that they accept.\textsuperscript{24} One side’s usage of ‘good‘ is probably causally tracked by a set of natural properties, which is constituted such that it excludes actions or states of affairs in which punishment that result in death is imposed. In contrast, the set of natural properties that causally track the opposing side’s usage of

\textsuperscript{22} Except of course that the staunch supporters of the opposing sides are certain that they are right, and that most people have deep convictions about the matter one way or the other. But not one participant or group involved in the debate can provide evidence similar to those that scientists offer for their theories that could vindicate their moral position.

\textsuperscript{23} In a debate like the one on capital punishment, the matter hinges not on which action or state of affairs is good but on which one is to be given greater evaluative weight. Participants in the debate are likely to share beliefs about the immeasurable value of a life, the irrevocability of meting out a death penalty, the merits of the existing judicial system, the fact that some crimes are worse than others, and so on. What they disagree about is comparative moral worth of something they both take to be morally valuable.

\textsuperscript{24} This is another way of saying that their use of moral terms is governed by different sets semantic norms. However, each set of semantic norms seems to include the rule that even in the event that moral terms and concepts are used differently a genuine disagreement can occur between speakers. It also seems that there is a more permissive threshold for disagreement about moral matters; despite a considerable difference in the semantic norms that govern speakers’ use of moral terms and substantive variance in moral principles disputants may still have a genuine moral disagreement.
‘good’ must be constituted such that it includes certain instantiations of the death penalty. It is possible to imagine the participants in this debate saying thus:

A: Imposing capital punishment violates my conception of what is morally right, which involves a duty to recognize the immeasurable value of human life that we can neither give nor take away from any individual. It is never right to impose capital punishment.

B: But I don’t see things the way you do. I think that under certain conditions imposing capital punishment can be morally right. My concept of morality entails that an act is right if it promotes the welfare of the greatest number and recognizes the demands of justice. Imposing capital punishment on those who commit heinous crimes clearly falls under this concept because it is the only acceptable form of redress in the circumstances and it deters further commission of such crimes.  

---

In order to show that there are problems with Frank Jackson’s moral naturalism, Denis Robinson discusses a similar type of moral disagreement:

"Argle: Assisted voluntary euthanasia violates my entire concept of what’s right, which holds human life sacrosanct. It is always wrong.

Bargle: Au contraire, it’s often right. According to my concept, what’s right is what maximizes well-being and minimizes suffering, so long as core moral values are respected. Autonomy is one of them. Coupled with the imperative to minimize suffering, it trumps the so-called sanctity of human life ["Moral Functionalism, Ethical Quasi-Relativism, and the Canberra Plan," Conceptual Analysis and Philosophical Naturalism (Cambridge: MIT Press, 2009), p. 318]."

Robinson thinks that there are some disagreements in which it is idle to ask which party is right from a neutral standpoint because no such standpoint exists. Any attempt to arrive at such a standpoint imports neutrality into one of the competing positions, thereby amounting to a petitio. But whereas Robinson thinks that this type of moral disagreement shows that moral functionalism implies quasi-relativism. Robinson rightly points out that this position is still a naturalistic form of moral realism. But what matters for present purposes is that the type of disagreement being discussed supports a new and interesting argument for antinaturalism. More on this in Chapter 4.
In a case like this, even someone who knows all descriptive facts about capital punishment may not know whether it is morally justifiable. The kind of convergence of moral opinion that could end the debate does not seem to be forthcoming. The difference between those who advocate and those who oppose capital punishment does not seem to depend on a difference in how they enumerate facts about matter under discussion. Rather, it rests on a difference in what they believe ought to be done given the facts. So much so that even someone who knows all natural facts may not be able to infer from what he knows something that could tell us which of the disputants got things right.

When we ask which analysis of goodness or sets of natural properties that causally track moral terms is correct we are asking in effect whether it is A or B who got things right. It is not only that we want to know if all things considered we ought to bring it about that the natural properties or analyses ought to be pursued, we want to know who is mistaken and who is right in using moral terms in the way that they do. Naturalists, like common folk, would say that A and B can’t both be right. But naturalists, unlike common folk, owe us an account of how we can determine which position is the right one. If such an explanation is not forthcoming, then moral naturalism is in trouble.

To mount a new argument against naturalism, a hypothetical moral agent who knows all the natural properties that constitute something but is still uncertain whether this something is good can be used to replace the tendency of competent speakers to ask the sort of questions that Moore raises. Being confronted with persistent disagreement between well-meaning and well-informed moral agents, who are also competent users of moral terms, could give someone who knows all natural facts a reason to think that
the wrongness of what moral agents disagree about may not be a natural, or
even a factual matter. What matters for this new antinaturalist argument is
the shift from talk of competent speakers asking open questions to talk about
fully informed moral agents who are uncertain whether something is good.
There seems to be no straightforward contradiction involved in knowing about
all the natural features of something and not being sure whether this thing is
morally good.

A fully informed moral agent could remain uncertain whether
something is good because he knows that competent speakers who are well-
formed about moral matters still disagree not only about what ought to be
done but also about which use of moral terms is right. Since knowledge of all
the natural features of something and competence with moral terms do not
 guarantee agreement about the moral goodness of that thing among moral
agents, the moral goodness things may not be a natural matter at all.

Although persistent moral disagreement is not ubiquitous the applica-
tion of complex moral notions like 'just punishment' to different sets of natural
facts ensure that the disagreement that can be gleaned from the moral twin
earth scenario is widespread. There are some moral matters about which
there seems to be a consensus, e.g. murder, slavery, torture for fun, etc.
Whether such a consensus is sufficient for establishing that a specific set of
natural properties causally regulate our use of moral terms as a matter of ob-
jective necessity is a separate question, especially since even well-informed
moral agents do not agree about what counts as murder, slavery or torture for
fun. It is possible that lack of agreement about the right use of moral terms
could undermine the apparent certitude brought about by our agreement on
murder, slavery, torture for fun, etc. Answering these questions could turn out
to be dependent on yet to be discovered empirical facts about what compe-
tent and well-meaning moral agents are disposed to desire and what it takes
for a moral disagreement like the one described to be resolved.

It still remains possible that moral goodness is natural, but then the
naturalist would also have to say that it is inaccessible to well-informed and
well-meaning moral agents, whom we nevertheless take to be competent with
moral terms, just which natural properties are objectively correlated with
moral terms. But of course it is absurd to say that the objectively correct use
of moral terms is inaccessible to competent speakers who are well-informed
about that natural features of a moral matter. So if it can be shown that this is
the only option that remains for the naturalist, then that would serve as a re-
ductio of their position.
Unlike Moore’s open question argument this new antinaturalist argument does not derive a metaphysical conclusion from a semantic premise. Instead, it derives a metaphysical conclusion from an epistemological premise. In the next four chapters this new ‘knowledge argument’ against moral naturalism will be discussed further. Here it was merely shown that since some moral disputes cast doubt on the possibility that our moral judgments are naturalistically justified, knowledge of natural facts are not sufficient for knowing whether something is morally good. As we shall see, this conclusion poses formidable problems for the moral naturalist.
Chapter 3

Introducing Jeeves:
someone who knows all natural facts
but does not know whether something is morally good

When he returned to Cambridge and to philosophy in 1929 after many years of self-imposed exile, Wittgenstein delivered a lecture on ethics.\(^1\) In this lecture he explicitly adopts some of Moore's ideas. "My subject, as you know is Ethics, and I will adopt the explanation of that term which Professor Moore has given in his book *Principia Ethica*."\(^2\) Despite this reference and Darwall, Gibbard and Railton's discussion of the similarity between


According to Klagge and Norman, LE was delivered to a society called 'The Heretics' on 17 November at the invitation of C.K. Ogden (p. 36). 1929 is the beginning of Wittgenstein's so-called middle-period (1929-1936), that time he was already inclined to repudiate what he wrote in the *Tractatus* but he had yet to write the views that were later published as the *Philosophical Investigations*. The editors of *The Philosophical Review* wrote in a note that the text of the lecture published in their journal "is a transcript of shorthand notes made by... Friedrich Waismann during and after conversations with Wittgenstein and Moritz Schlick in 1929 and 1930 (p. 3)." Wittgenstein repeated the main lines of his lecture to Waisman and Schlick between December 1929 and January 1930 [Ray Monk, *Ludwig Wittgenstein: the Duty of Genius* (London: Jonathan Cape, 1990), p. 282]. G.H. von Wright's catalogue of Wittgenstein's papers lists two existing version of LE: a 22-page handwritten manuscript (MS 139a) and a 10-page typescript (TS 207). Von Wright claims that "139a differs in some interesting respects from the typescript (207) from which the lecture was posthumously printed in the *Philosophical Review*. [Georg Henrik von Wright, "The Wittgenstein Papers," *Philosophical Occasions 1912-1951*, pp. 496-7]

Brian McGuinness speculates that LE is probably based on a German text which is now lost. [*Wittgenstein and the Vienna Circle: Conversations Recorded by Friedrich Waismann* (New York: Barnes & Noble, 1979), p. 92 n. 60.]

\(^2\) LE p. 4
Wittgenstein’s and Moore’s views on ethics, the parallel between their views have neither been explored nor appreciated much. Some even go so far as to say that Wittgenstein’s lecture is of no significance whatsoever. But apart from explicitly referring to Moore in his lecture, Wittgenstein also presents an argument against moral naturalism reminiscent of the open question argument. Wittgenstein's argument, which is based on the possible moral ignorance of someone who knows all natural facts, bears a striking similarity to Moore’s, which relies on competent speakers who persist in asking certain questions.

This new antinaturalist argument rests on the idea that if someone who knows all natural facts would not know whether something is morally good, then moral goodness must not be a natural matter. I will call this argument the world book argument (WBA) and the omniscient person in it Jeeves, after the knowledgeable butler in P.G. Wodehouse’s writings. In this chapter it will be shown that the world book argument presents a formidable


“Wittgenstein’s Lecture on Ethics is undoubtedly one of the shoddiest things ever written on the subject (p. 118).”

“Wittgenstein’s ‘Lecture on Ethics’ is of no worth whatever for ethical inquiry, and ... the manner of philosophizing which it exhibits is despicable (p. 127).”

5 This is something that Darwall, Gibbard and Railton, or even David Wiggins who pointed out the Wittgenstein reference to them, seem to have missed.

6 Jeeves has long been associated with being a helpful know-it-all, as evidenced by the internet search engine named after him (askjeeves.com), one of the precursors of the now ubiquitous google. Jeeves, like Wittgenstein’s omniscient person and the ideal advisor or observer in some ethical theories, is someone we typically defer to regarding correct moral judgment. But at the same time, Jeeves and other similar characters are constructed as an aid to limited and fallible moral agents. Archangels, omniscient persons, ideal observers and fully rational versions of ourselves are all aides to whom we defer, so naming such a character Jeeves is fitting.
challenge to moral naturalism because it is immune from some of the standard objections to Moore's argument and because it shares some of the advantages of another well-known argument: Jackson's knowledge argument against physicalism. The world book argument embodies the new and more promising argument against moral naturalism that began to emerge at the end of the previous chapter.

Wittgenstein's Argument Against Moral Naturalism

It is unclear whether Wittgenstein intended his discussion of an omniscient person's ignorance about ethics to be a basis for a new antinaturalist argument similar to Moore's. There is insufficient evidence in the published writings of both philosophers to indicate anything definite on this point. Nevertheless, it is obvious that in the following passage Wittgenstein expresses some interesting ideas that bolster the new argument against moral naturalism that emerged at the end of Chapter 2.

Now what I wish to contend is that, although all judgments of relative value can be shown to be mere statements of facts, no statement of fact can ever be or imply, a judgment of absolute value. Suppose one of you were an omniscient person and therefore knew all the movements of all the bodies in the world dead or alive and that he also knew all the states of mind of all human beings that ever lived, and suppose this man wrote all he knew in a big book, then this book would contain the whole description of the world; and what I want to say is that this book would contain nothing that would be called an ethical judgment or anything that would logically imply such a judgment. It would of course contain all relative judgments of value and all true scientific propositions and in fact all true propositions that can be made. But all facts described would, as it were, stand on the same level and in the same way all propositions stand on the same level.\(^7\)

Wittgenstein thinks that someone fully informed about the physical and mental features of the world will not know whether something is good. This implies that someone who remains unsure whether killing another human being

\(^7\) LE p. 6
is good does not necessarily remain uncertain because of ordinary ignorance, i.e. lack of information about the natural features of such an act. Wittgenstein’s omniscient person scenario is supported by the traditional distinction between instrumental and intrinsic value; Wittgenstein calls the former relative or trivial and the latter absolute. He rightly points out that statements that express the instrumental value of something can always be reformulated as a statement of fact. For example, when we say that someone is a good police officer we mean that she enforces the law according to established procedure with a certain degree of efficiency. Here enforcement of the law is the purpose relative to which a police officer is evaluated. The value involved in the relative or instrumental sense of ‘good’ is a function of a means-ends relation between a predetermined purpose and something that brings it about. In contrast, ethics is concerned with what is good ‘in itself,’ like Moore, Wittgenstein thinks that this is the kind of value that ethical expressions convey. Intrinsic value is usually described as something that is pursued for its own sake.

For instance, we think that the enforcement of the law is good because it allows a large number of people to live happy and productive lives; the latter is a good candidate for something intrinsically valuable. We think that the collective happiness of a great number of people is worth pursuing for its own sake; it is not worthwhile merely because it serves another purpose. If someone asks us why we think that the collective happiness of a great number of people is good we might describe the situation in more detail and say that it is a situation in which many people are pleased. We might then ask our interlocutor to imagine herself happy or pleased and ask her if she thinks that some things are good because it contributes to bringing about such a state. If she gives an affirmative answer then we might add that the happiness of a great number of people is that from which many other things
derive their goodness. So it is the collective happiness of a great number of people from which the goodness of other things derive; such a situation is non-derivatively good. If our interlocutor persists in asking how we are sure about the goodness of the happiness of a great number of people we might wonder what we could do to convince her that we are right. She might persist in asking questions because she does not see that valuing her own happiness automatically gives her a reason to value the happiness of a great number of people, whether it includes her or not. This development merely shows that valuing something because of its intrinsic moral worth is different from valuing something for prudential reasons. In any case, we think that we are right about the non-derivative moral worth of the happiness of a great number of people, but how do we know this? Does our interlocutor persist in asking questions because there is some piece of information about the natural world that she lacks?

Wittgenstein favors a negative answer to this question. The quote from his lecture above implies that the following two propositions ‘stand on the same level:’ (1) A great number of people are happy; and (2) A great number of people are suffering. He thinks that a person omniscient about all the natural facts could be ignorant about the important ethical difference between (1) and (2). In writing a world book a person omniscient about all the natural facts will describe the world neutrally, his book would contain either (1)

\[\text{8 There is a certain sense in which evaluative statements like 'A situation in which a great number of people are happy is intrinsically good' cannot be proved. In this way, even those evaluative statements about which virtually everyone agrees have the same status as the existence of the external world, the existence of other minds and other common sense propositions. They are not indubitable. However, once it is granted that knowledge of natural facts is possible, even if not all natural facts are cognitively accessible, doubt about non-evaluative common sense propositions disappear. Wittgenstein's point about ethical judgments is that even if we grant that someone knows all natural facts, it is still possible to doubt whether such a person knows that the greatest happiness of the greatest number is intrinsically good.} \]
or (2) but it would not contain the judgment that (1) is good and (2) is bad, and the writer of the world book might not think that he must do something to bring about (1) and prevent (2) from happening or continuing to happen. Wittgenstein goes so far as to say that in this world book murder “will be on exactly the same level as any other event, for instance the falling of a stone.”

This is a provocative way of putting the point about the complete absence of evaluative statements in the omniscient person’s description of the world. It is certainly possible to describe the world in value-neutral terms. But Wittgenstein makes a stronger claim: a complete value-neutral description of the world, i.e. a world book, would not contain anything that implies a moral judgment. So when someone omniscient about all the natural facts writes a world book, this book would include the fact that some moral judgments are accepted by virtually everyone, but it would not include any statement that would count as a moral fact.

Since the omniscient person’s world book does not contain any moral facts, it follows that the omniscient person is morally ignorant. After all, such a person cannot refrain from making other judgments because of what he knows, e.g. naturalistic judgments that are necessary for knowing exactly how the human brain works or socioeconomic judgments that are necessary for fully understanding what events created the recent economic recession. If someone omniscient about all natural facts does not make moral judgments, then this must mean that judgments made in the moral domain do not stand on the same level as naturalistic, socioeconomic, and other similar conclusions that can be drawn from natural facts. Moral judgments then must be

---

9 LE p. 6
different from other types of judgments.\textsuperscript{10} Wittgenstein thinks that the absence of any ethical statements in the omniscient person’s world book is a sufficient basis for saying that the goodness must either be ineffable or nonexistent.

It is noteworthy that in his lecture Wittgenstein ends with a disjunctive conclusion: ethics is either supernatural or nonexistent.\textsuperscript{11} He even goes so far as to say that he is inclined to believe that ethics falls within the domain of the mystical.\textsuperscript{12} Since Wittgenstein does not say more about what he clearly takes to be ineffable it is difficult to say what the claim that ethics is supernatural amounts to. It is worth noting, however, that even the claim that ethical properties are supernatural or nonnatural is prone to the objection posed by the type of question raised by Moore in his open question argument.\textsuperscript{13} Even if good means ‘is approved by God,’ it is possible to ask of something, which is approved by God, whether it is good. Although Wittgenstein and the Ancient Greeks may not have had the same deity in mind, Plato’s \textit{Euthyphro} demonstrates very well that defining piety or goodness in terms of what is

\textsuperscript{10} It is possible that not only moral judgments but also all evaluative judgments differ from natural or descriptive ones.

\textsuperscript{11} “Ethics, if it is anything at all, is supernatural (LE p. 7).”

\textsuperscript{12} “My whole tendency and I believe the tendency of all men who ever tried to write or talk Ethics or Religion was to run against the boundaries of language. This running against the boundaries of our cage is perfectly, absolutely hopeless. Ethics so far as it springs from the desire to say something about the ultimate meaning of life, the absolute good, the absolute valuable, can be no science. What it says does not add to our knowledge in any sense. But it is a document of tendency in the human mind which I personally cannot help respecting deeply and I would not for my life ridicule it (LE pp. 11-12).”

\textsuperscript{13} Moore himself did not see that the strategy he used to undermine that moral naturalist’s position can be used against moral nonnaturalism. This is probably why some philosophers call Moore’s deployment of the open question argument accident-prone (Darwall, Gibbard and Railton, p. 115).
loved by the gods gives rise to a formidable problem. If something is good because it is loved by the gods then this would make goodness arbitrary but if the gods love something because it is good then we still do not know what makes things good. Perhaps it is because of the latter possibility that we could still ask whether something approved by the gods is good. Whatever the reason, it is ineffectual to define goodness in terms of what is loved or approved by the gods. That is why notwithstanding Wittgenstein’s predilection for mysticism, an antinaturalist conclusion with antirealist affinities fits well with the argument that he gives in his lecture. The world book argument can be stated thus:

WB1 Someone who knows all natural facts would fail to know whether something is morally good.

WB2 If someone who knows all natural facts does not know whether something is morally good, then moral naturalism must be false.

WB3 So moral naturalism is false.

Stated in this way the world book argument has obvious parallels with Moore’s argument against moral naturalism. But instead of deriving a metaphysical conclusion from semantic premises, this new antinaturalist argument uses an epistemological premise to arrive at a metaphysical conclusion. If a person omniscient about all the natural facts does not know whether something is morally good, then knowing whether something is morally good must not be a natural matter. In the previous chapter we have already begun to see that stating the antinaturalist position in this way has significant advantages, here we will see the advantages of the world book argument in greater detail.

---

The world book argument and the open question argument

Stipulating omniscience about natural facts alone cannot propel an argument against moral naturalism; in itself it merely clarifies what is at stake in a debate in which one side claims that certain things exist while the other denies their very existence.\(^{15}\) This is why it is necessary to explain what it means for Jeeves to be omniscient about all the natural facts but ignorant about moral goodness and why he is the way that he is. In the previous section it was shown that an omniscient person would not necessarily make ethical judgments, i.e. Jeeves can write a complete description of the world without making any moral judgments. In this section we will begin to see why such a possibility implies that naturalistic forms of moral realism are in trouble.

The reason Jeeves’s failure to make moral judgments implies that moral properties are not natural has something to do with the type of moral disagreement discussed in the last section of the previous chapter.\(^{16}\) In this

\(^{15}\) The idea of omniscience alone is neutral. Using the idea of knowing all natural facts alone can only go so far, it is a tool that both realists and antirealists about a certain domain could use to clarify their position.

“The idea of an omniscient God is often appealed to by philosophers who have no belief, or even interest, in the existence of such a being. An omniscient God is a useful expository device for realists who believe that there are true propositions which we human beings may never be in a position to assert with good warrant. They may be truths about the remote past, about remote corners of the universe, about Schrödinger’s cat, or about the secrets of the human heart; the idea is that, while such matters may be impenetrable to us, we can think of God as knowing the truth of the matter.

Anti-realists, by contrast, may invoke the idea of an omniscient God in almost the opposite way. Consider, for example, the question ‘If John Kennedy hadn’t been assassinated, would George Bush still have been elected president of the United States in 1988?’ A philosopher is anti-realist about counterfactuals might say that there is here no truth of the matter, nothing that even an omniscient God could know.” [Murray MacBeath and Paul Helm. “Omniscience and Eternity,” Proceedings of the Aristotelian Society, Supplementary Volumes 63 (1989): p. 55.]

\(^{16}\) See pp. 28-31.
type of moral disagreement the disputants agree about the natural features of
the matter under discussion, but disagree about what moral conclusions to
draw from these natural features. It was previously shown that if Moore's
argument against moral naturalism had relied on this type of disagreement
rather than on the questions of the sort he posed about naturalistic analyses
of moral goodness, his argument would have fared better. Here it will be
shown that the world book argument is better than the open question
argument precisely because it makes use of the relevant type of moral
disagreement to make a case against moral naturalism.

Ordinary use of moral terms and concepts has the appearance of
being descriptive; statements that feature moral predicates resemble
statements that feature natural property attributions. Moreover, the moral
goodness of something co-varies with the natural features of that thing.
These considerations partly support the moral naturalist's claim that correct
usage of moral terms and concepts are causally tracked by a set of natural
properties. But since moral agents who have access to the same natural
facts about the moral matter under discussion can arrive at different moral
conclusions about it, doubts about the usefulness of the notion of 'causal
tracking' for establishing the claim that moral properties are natural arise.
Each participant in the relevant moral dispute would support her conclusion
by citing natural facts and appealing to the normative moral principles to
which she subscribes, which each could in turn show to be congruent with
natural facts. But since more than one set of normative moral principles are
congruent with natural facts we have to ask how, if at all, we can know for
sure that this or that set of normative moral principles is justified. Without
finding some neutral standpoint from which to arbitrate such a disagreement
by showing just which moral principles are correct, the whole notion of 'causal
tracking’ begins to look dubious. That there are at least two different ways of
drawing moral conclusions from the same set of natural facts implies that
drawing moral conclusions from natural facts is not a matter of making an
objective inference.

If moral naturalists are right, i.e., if it is true that moral properties are
natural properties, then surely sincere and well-informed moral agents must
arrive at similar moral conclusions given that they agree about the natural
features of some moral matter. And someone who knows all the natural
features of some moral matter must know all about its moral properties also,
as a matter of necessity. But as we have seen previously sincere and well-
formed moral agents do not arrive at similar moral conclusions about say,
capital punishment, despite agreement about the natural features of the moral
matter under discussion. Such an impasse makes it unclear whether moral
principles can be justified. Jeeves can only know whether something is
morally good if he is able to derive the evaluative from the descriptive in the
same way that he derives higher order physical laws from complete natural
information about moving bodies. That is, uncontroversially and
demonstrable to the well-meaning and well-informed. The problem isn't so
much that it is impossible to derive the evaluative from the descriptive. After
all, the disputants in the relevant type of moral disagreement were able to do
it. The problem is that it is difficult to say which, if any of them, is right. Each
side is justified by their own lights. That is why Jeeves is unable to say which
of them is right, which is ultimately why he does not know whether something
is morally good.

If it is possible to find justified moral principles, then bringing them to
bear on the descriptive would yield the right moral judgment. But since there
is no uncontrovertial way to establish that some specific set of normative
moral principles are justified it is plausible to assume that Jeeves could know what he does and still fail to know whether something is morally good because he does not know which moral principles, if any, are justified.

A dispute between moral agents similar to A and B's disagreement about capital punishment described previously poses a problem for Jeeves. He knows about all such disagreements. And it is possible that he is not sure whether he can successfully arbitrate these disagreements by solely appealing to the natural facts that he knows. Since statements that feature an attribution of a natural property do not provide a reason to do something while statements that feature attributions of moral goodness do, it is not clear how descriptive premises can yield a normative conclusion without some illicit attempt to import evaluative concepts onto natural facts. Here we see the core of the problem that the relevant type of moral disagreement poses. Each disputant in such a disagreement brings the moral principles to which she subscribes to bear on the natural features of the matter under discussion. So the moral conclusion that each participant in the debate reaches is determined by which natural facts are given greater moral weight by the moral principles to which she subscribes. In the disagreement between A and B, the latter will look at the effects of imposing capital punishment on the welfare of the greatest number while the former will look at what type of action the imposition of capital punishment is and the reasons why such actions must be forbidden. Because of the difference in their evaluative priorities, A and B arrive at different conclusions about the moral worth of capital punishment.

Jeeves's knowledge of the different moral valuations of the same moral matter puts him in a predicament. Those who argue for and against capital punishment support their position by appealing to the natural features

---

17 See p. 31.
of the moral matter in question and they talk and think like their judgment automatically follow from the natural features of capital punishment. There would be no problem if these putative judgments are all of a piece. But the fact that these judgments are opposed lead Jeeves to think that uncertain whether the mere fact that capital punishment has the natural features that it does not give him a reason to make some moral judgment about it, much less that some judgment on the matter is justified.

The first premise of the world book argument is more effective than the analogous premise of Moore’s argument. Wittgenstein’s omniscient person scenario presents a challenge not only to those naturalists who propose an analytic or a priori equivalence between natural properties and moral ones. Even naturalists who favor synthetic or a posteriori analyses of moral goodness need to address the challenge presented by Jeeves. His knowledge of all natural facts entails that if he is morally ignorant, then knowledge of moral goodness must be distinct from knowledge of all the natural features of all morally good things in a way that is detrimental to moral naturalism. If knowledge of all natural facts does not entail moral knowledge then the latter might not turn out to be knowledge at all. Since Jeeves, as he has been described, presents a coherent idea, this idea could potentially establish very clearly that moral properties are not natural properties. At the same time, the first premise of the world book argument is more controversial than the first premise of the open question argument. So a lot more needs to be said about why fallible omniscience implies antinaturalism.

---

18 So much so that some naturalists might be inclined to say that like OQ1, WB1 merely amounts to a petitio. This issue will be taken up in the next chapter. It is worth noting here though that the version of a premise or an argument that is more controversial is also more effective.
Unlike its analogue in the open question argument,\textsuperscript{19} the second premise of the world book argument is not prone to those objections posed against Moore’s argument. The first premise of the world book argument is effective against non-analytic forms of moral naturalism. Jeeves’s omniscience entails that he knows about all synthetic natural identity claims. His possible moral ignorance cannot be explained away by making an appeal to new moral semantics. Because knowing whether something is morally good comes apart from knowing the complete natural description of all morally good things, there must be an insurmountable distinction between moral properties and natural ones. And because the world book argument does not depend on a dubious assumption about the philosophical significance of the question ‘But is it good?’ it cannot be dismissed as easily as Moore’s argument. It remains to be seen whether the second premise of the new argument also fares better than its analogue in the open question argument. It is to this issue that we must now turn.

\textbf{The world book argument and the knowledge argument (KA)\textsuperscript{20}}

There are some striking similarities between the world book argument and Jackson’s knowledge argument.\textsuperscript{21} Both arguments are propelled by the same strategy: presenting a vivid description of someone who knows all the facts relevant to some domain but still ignorant about it in an important way.

\textsuperscript{19} OQ2: If for any natural property it is an open question whether something which has this property is morally good, then moral naturalism must be false.


\textsuperscript{21} For a discussion of the parallels between Moore’s OQA and Jackson’s KA see Michael Pelczar, “The Knowledge Argument, the Open Question Argument, and the Moral Problem,” \textit{Synthese} 171 (2009): 25-45. My discussion of the parallels between WBA and KA here was set off by ideas in this paper.
Whereas the world book argument has Jeeves, the knowledge argument has Mary:

Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes or the sky, and use terms like 'red,' 'blue,' and so on. She discovers, for example just which wavelength combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that result in the uttering of the sentence 'The sky is blue.' (It can hardly be denied that it is in principle possible to obtain all this physical information from black and white television, otherwise the Open University would of necessity need to use colour television)

What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it.22

Of course Wittgenstein’s lecture on ethics was written many years before Jackson’s argument was published. And there seems to be no reason to think that Jackson developed the knowledge argument by taking some of Wittgenstein’s ideas. But it is noteworthy that a precursor of the knowledge argument can be found in the writings of one of Wittgenstein’s contemporaries. C.D. Broad presents a character similar to Jeeves in order to show that knowledge of natural facts excludes direct acquaintance with properties that are emergent from these facts. In The Mind and Its Place in Nature there is a discussion of a mathematical archangel who knows all the microphysical properties of nitrogen, hydrogen and the gaseous compound which these can form but does not know the smell of ammonia.23

A variation on the same argumentative strategy is expressed in the following question: “Could a Martian, entirely without sentiments of compassion or piety, know about what is going

---

22 Jackson, “Epiphenomenal Qualia,” p. 130

on during a commemoration of an armistice?" Jackson improves on the scenario described by Broad, Feigl and others by providing more details about the situation of the well-informed subject before and after she gets directly acquainted with those properties that she is putatively ignorant about. Mary and her black in her white room presents a vivid picture of how someone can be fully informed about something and yet still ignorant about some aspect of it. Jeeves is much like Mary, the archangel and the Martian; they are all ideal and fully informed yet deficient in some way.

To compare the knowledge argument with the world book argument in a straightforward way the following version of Jackson's argument will be used here:

K1 Someone who knows all physical facts about color vision would not know what it's like to see red.

K2 If someone who knows all physical facts about color vision does not know what it's like to see red, then physicalism about conscious experience must be false.

K3 So physicalism about conscious experience is false.

The most crucial part of the knowledge argument is its second premise. It presents the problem that Mary poses for physicalists in a clear and intuitive way. Because of the coherence of the Mary scenario a clear cut division is

24 Herbert Feigl asks this question in an article on the mind-body problem.

“For the sake of the argument, we assume complete physical ... predictability and explainability of the behavior of humans equipped with vision, a sense of humour and sentiments of piety. The Martian could then predict all responses, including the linguistic utterances of the earthlings in the situations which involve their visual perceptions, their laughter about jokes, or their (solemn) behavior at the commemoration. But ex hypothesi, the Martian would be lacking completely in the sore of imagery and empathy which depends on familiarity (direct acquaintance) with the kinds of qualia to be imaged or empathized.” [“The ‘Mental’ and the ‘Physical,” Minnesota Studies in the Philosophy of Science 2 (1958): 431.]
established between knowledge about color vision and the experience of what it is like to see red. Since it is apparent that bridging such a division presents formidable problems there is reason to think that knowing what it’s like to see red cannot be reduced to knowing all the facts about seeing red. Stuart Brock and Edwin Mares take the knowledge argument to be a paradigm example of a strategy that they call an argument from an attitude problem. The argument relies on our conflicting attitudes towards physical information about color vision and what it is like to see red. Furthermore, this discrepancy in our attitudes towards two closely related domains is taken as a basis for saying that what it is like to see red cannot be reduced to physical information about color vision. In other words, the incongruity of our attitudes towards the reduced properties and the proposed reductive base proves that the attempt at reduction has failed. Other standard features of an attitude problem include citing special reasons why one domain is not reducible to another and stipulating that the subject involved is not suffering from some kind of debilitating ignorance or cognitive malfunction.

In the knowledge argument it is implied that whereas the chemistry and the biology of color vision can be reduced to physical facts, the phenomenology of it cannot. Like the knowledge argument, the world book argument features a fully informed subject who is not suffering from debilitating ignorance or cognitive malfunction. Likewise, Wittgenstein’s antinaturalist argument is built around conflicting attitudes towards two related domains: knowledge about the natural features of something and knowledge of that thing’s moral worth. But unlike the knowledge argument, which establishes

25 Stuart Brock and Edwin Mares, Realism and Anti-Realism (Stocksfield: Acumen, 2007) 24-7.

26 Ibid.
the distinct metaphysical existence of what it is like to see red, the world book argument arrives at the conclusion that there are no moral facts. This difference between the conclusions of the two arguments might have something to do with the distinct features of the subject matter of each. Whereas in philosophy of mind the mind itself cannot be coherently doubted, in metaethics one of the central concerns is whether there really are such things as moral properties in the same way that there are natural ones.

The second premise of the world book argument can be supported by drawing a parallel with the knowledge argument. Naturalists think that moral facts, or at least moral properties, are already covered when we talk of all physical and mental facts that constitute the world. But naturalists are usually silent about whether these moral properties would appear to us or to Jeeves in the same guise as the natural ones do, especially if he does not participate in our moral practices like Heigl’s Martian. Since there is a problem involved in drawing moral conclusions from purely descriptive premises it is reasonable to suppose that even if descriptivist naturalism were true moral properties could only become manifest in a distinctive guise which is not easy to square with a mere enumeration of all the relevant natural features of something. That is, moral properties which are fully captured by a natural description of something that has them may still turn out to be recognizable as such only under a certain distinctive guise. If this is the case the naturalists must explain why moral properties have this distinctive guise at all. But before the plausibility of such an explanation can be settled it remains a distinct pos-

---

27 René Descartes famously discovered this when he tried to doubt everything. *Meditations on First Philosophy* (New York: Cambridge University Press, 1996), 16-23.

28 This old fact-new guise strategy against the world book argument will be the main topic of Chapter 6.
sibility that moral properties are not natural. Of course it remains to be specified further what exactly it means for a property to be natural or otherwise. But since some descriptive properties that are ordinarily taken to be natural, e.g. psychological and economic ones, do not generate as much difficulty and disagreement as moral properties, it is probable that there is something about moral properties that sets them apart.

There are some other domains in which conceiving of someone like Jeeves or Mary implies that the properties associated with the domain in question are not natural in the same sense that economic and psychological ones are. The second premise of the world book argument can also be supported by drawing an analogy between ethics on the one hand, and comedy and language on the other. We can conceive of someone who knows all comedic facts, including the biological, sociological, and even literary aspects of what human beings find funny, but still does not know whether some new sketch is funny. Call this comedically challenged hypothetical being, Jean. Like Jeeves, Jean can write a complete natural description of all comedic facts. But in the same way that Jeeves’s world book would not include any moral judgment, Jean’s book of comedy would not include any comedic appraisal.

Crispin Wright offers an account of the independence of what is funny from our reaction to it that could be relevant to understanding why someone like Jean is conceivable:

When I claim that something is funny, I’m not simply reporting my own reaction to it; for I can readily conceive that my reaction might be wrong—insensitive or misplaced. Nor am I reporting a majority reaction to it, or conjecturing the direction it might take; for the same applies—the comic sensibilities of the majority may be blunted, "off the wall" or debased. Re-
garding something as funny incorporates a judgment about the *fittingness* of the comic response.\textsuperscript{29}

Jean's uncertainty can be explained by a worry about distorting factors that could affect her reaction to something. Our comedic response is affected by our biases, among other things, and she knows this. The comedic properties of something is not up to us; just because we laugh at something or think that it is funny does not mean that it is appropriate to laugh at it. Some corrective measures are even adopted when we think it is possible that what we find funny could turn out to be offensive or mean. Perhaps Jean is a logical possibility because someone fully informed about the comedic can have a debased sense of humor. Knowing all comedic facts is certainly compatible with being bigoted about what counts as funny. Knowing all comedic facts is also compatible with not knowing what we ought to find funny.

But these considerations are not the only possible explanations for the conceivability of someone like Jean. It is possible that what such a person knows is not sufficient to determine whether that thing is funny. That is, what Jean knows does not determine the comedic properties of something one way or another. This implies that knowledge whether the new sketch is funny cannot be inferred from knowing all the relevant facts, or even relying on an individual's or a population's instinct or gut reaction to that thing. But since there is no independent access to the appropriateness of a comedic response apart from reliance on people's responses and engaging upon discussion and reflection on what we find funny, it seems reasonable to take an antirealistic line on comedic properties. Knowing whether something is funny is different from knowing whether something is red because there is no independent access to the former. This difference might be part of the reason an antirealist

construal of comedic properties seems to be warranted while an antirealist construal of surface reflectance properties associated with color vision is not. Moral properties are like comedic properties in that there is no independent access to them apart from reliance on our intuition about cases and a similar, albeit more systematic, discussion and reflection upon these. So it is also possible to construe moral properties in an antirealist way. What this analogy between ethics and comedy shows is that a purely descriptive account of the world may not be able to generate an account of what we ought to find funny or what we ought to take to be morally good or bad because, as it has been shown in the previous section, it is difficult if not impossible to settle disputes about judgments like these by showing which of them is justified.

In the same way that we can conceive of Jeeves and Jean we can conceive of someone who knows all linguistic facts, including all psychological, neurological and evolutionary facts about language, but does not know how to speak good conversational English. Call this hypothetical being Jane. Jane knows all the rules of English grammar; she is also familiar with a wide variety of words. Perhaps we can even say that she knows everything written in the Unabridged Oxford English Dictionary. Besides this, she knows everything about both the literary and scientific aspects of every language, including English. Jean can write all that she knows in a book and in this book will contain all linguistic facts. But Jean’s book on language will not help anyone who, like her, needs help with conversational English. Despite knowing all that she does Jane can be stumped when she finds herself chatting with ordinary English speakers. Jane may not know what would be appropriate to say in particular conversational situations such that talking to her could turn out to be like trying to communicate with a Martian. Jane does not know what she ought to say in particular conversational situations in the same way that Jean
does not know whether she ought to find a new sketch funny. What one ought to say when conversing with others and what one ought to find funny are not merely a function of our reactions to it. Making some particular remark may be appropriate but it may also be rude or mean, so that one ought not to say it. Jane could fail to know what to say in ordinary conversational situations because she cannot understand how, if at all, what one ought to say follows from what she already knows.

Like the capacity for moral judgments, the capacity for language is shared by virtually all humans. But having this shared capacity and an established body of shared judgments do not necessarily imply that there are natural facts objectively correlated with moral goodness. This may be because our tendency to think of some things as morally good or bad is underdetermined by natural facts, including facts about our natural dispositions. Whatever the reason, the second premise of the world book argument shows us that there are important aspects of the ethical that the descriptive cannot accommodate and that the reason for this could very well be that moral naturalism is false.


De Waal's hypothesis is not being endorsed here. Indeed, it is difficult to say what an endorsement of an empirical hypothesis from philosophical considerations amount to. De Waal’s hypothesis is merely being presented as one among many possibilities that would flout the moral naturalist's expectations. More on this on p. 59 and following.
Chapter 4

Taking Jeeves Seriously:

would someone who knows all natural facts
really fail to know whether something is good?

When juxtaposed with Moore’s argument against moral naturalism and Jackson’s argument against physicalism, it is surprising that the world book argument is not more well-known or widely discussed. Coupled with the proliferation of full information accounts of moral goodness, the relative obscurity of Wittgenstein’s argument against moral naturalism seems even more astounding. It is possible that the world book argument is not more well-known or widely discussed because philosophers find it difficult to take the notion of fallible omniscience seriously. What lies behind this difficulty is not clear because people don’t usually give reasons for not talking about what they don’t discuss. In this chapter, the first premise of the new antinaturalist argument will be explained and defended in order to facilitate a discussion of the world book argument.

The Notion of Fallible Omniscience

The first premise of the argument against moral naturalism from Wittgenstein’s lecture on ethics expresses the logical possibility of knowing all natural facts without knowing whether something is morally good. Jeeves as he has been described in the previous chapter embodies this logical possibility. Such a possibility does not depend on arcane ideas. Even common folk can recognize that there is a difference between knowing whether something is morally good and knowing all the natural features of that thing. We ordinarily think that someone well-informed about natural facts
is different from someone who knows the difference between right and wrong. It is neither contradictory nor incoherent to say of someone that she is intelligent and well-informed but muddled about the difference between good and bad; characters like this abound in fiction and in real life. Jeeves could merely be an extreme version of such a character. He could be a natural byproduct of our ordinary understanding of knowing whether something is morally good, knowledge that can come apart from knowledge of the natural features of that thing.

Attributing knowledge of all natural facts to someone is not the same thing as attributing moral infallibility to that person. We see this best when philosophers talk about God. When St. Anselm describes God as ‘that than which nothing greater can be conceived,’ he is in effect saying many things about God. As the most perfect existent, much is attributed to God: he is all-knowing, all-powerful, eternally happy and supremely good. But not only that. The most singular attribute of God is his self-sufficient existence; only God exists of and by reason of Himself. It is from this last attribute that philosophers argue for the perfection of God. In contrast, Jeeves merely knows all natural facts. He is perfect only in terms of knowledge of the natural world. Jeeve’s knowledge of all natural facts is certainly compatible with moral infallibility. But his omniscience is also compatible with moral perversion, and with neutrality on moral matters. This is because after omniscience about natural facts is attributed to Jeeves more has to be said about him before we can say anything about his moral knowledge or lack thereof. To say that Jeeves knows all natural facts is not the same thing as saying that he knows all moral facts. Because of this difference it is possible for him to possess the former without possessing the latter.
Wittgenstein has it that a person who knows all natural facts could give a comprehensive description of the world without making any moral judgments. The world book would contain reports about all moral judgments that were ever made, including an account of how these were made. But to all this Jeeves would not add anything that might be construed as his own moral judgment; he would only list all facts, including facts about our moral practices. We could add that the world book is written in the way that Wittgenstein suggests because Jeeves is a dispassionate alien who wants to write an accurate account of what our planet is like. The material point is that a complete description of the natural world can be written without any morally evaluative statements.

Jeeves could even note that virtually everyone on our planet thinks that slavery and torture for fun, among other things, are morally wrong. Perhaps he would even add that these and other widely held moral judgments are made through the proper functioning of highly evolved neural mechanisms and that such judgments and their accompaniments aid in the survival and flourishing of our species. He would also note that more sophisticated moral distinctions, which are learned by moral agents as they mature, contribute to the highly complex social structures and institutions that preserve and enforce a code of behavior that human beings need to subsist for as long as possible. But Jeeves stops at this. He neither agrees nor disagrees with the moral judgments that we make; he neither subscribes to nor rejects our morals. This description of Jeeves does not involve any obvious contradiction; conceiving of someone omniscient about natural facts yet morally fallible does not amount to imagining a female brother or a married bachelor.
What then makes it difficult for philosophers to take Jeeves seriously? Perhaps it is simply the obviousness of the fact that morality is part of the natural world. So much so that someone who knows everything there is to know about the natural world ought to know about morality also. After all, it cannot be denied that all humans engage in the practice of making moral judgments. Such a practice has created a variety of moral codes, which are articulated in moral language. Hence moral judgments, moral rules, and moral predicates, etc., are just some of the natural phenomena studied by psychologists, neuroscientists, anthropologists, ethnographers, linguists, and so on. But it is one thing to know all that these specialists know about moral phenomena and quite another to know whether something is morally good. Fallible omniscience is not presented as a denial of the fact that moral phenomena are one of the things that exist in the natural world. Rather, it is an expression of the possibility that a certain type of explanation for moral phenomena may not exist. Just as a moral naturalist is not merely saying that these moral phenomena exist, the scenario presented in the first premise of the world book argument does not merely amount to saying that such phenomena are nonexistent. What the new antinaturalist argument amounts to saying is that contrary to what the naturalist expects, there could be nothing that we can derive from natural facts regarding morality apart from a genealogical explanation for the practice of making moral judgments, a detailed comparative analysis of different moral systems, and a functional-evolutionary account of the neural mechanisms that are implicated in moral judgment.

That it is possible to imagine someone like Jeeves presents a problem for moral naturalists because a reasonable concomitant of the view that moral properties are natural is the expectation that someone who knows everything
there is to know about natural properties would be as fully informed about moral properties also. The first premise of the world book argument is not a denial of the coherence of this expectation. If moral naturalism is true, then it is reasonable to expect that someone who knows all natural facts would be morally infallible. Fallible omniscience is an articulation of an imaginative possibility that flouts the naturalist’s expectation. If substantial philosophical reasons can be given for taking the possibility that Jeeves presents seriously, then it will begin to seem that moral naturalism cannot be true. The world book argument is built around the idea that if the naturalist’s expectations prove to be ill-founded then we have at least one good reason to believe that moral properties are not natural. The plausibility and coherence of fallible omniscience is a counterexample to moral naturalism, much like a zombie is a counterexample to physicalism about conscious experience.¹

A Naturalist’s Account of Omniscience

A proponent of the world book argument has it that someone who knows all natural facts but does not know whether something is morally good is a counterexample to moral naturalism. But moral naturalists think that someone who knows all natural facts would always know whether something is morally good. When naturalists conceive of someone like Jeeves they

¹ Jeeves is like phenomenal zombies in philosophy of mind, beings exactly like us in all physical respects but lacking conscious experience. Briefly, there is nothing it is like to be a zombie. Zombies are important because as a logical possibility they count as a counterexample to physicalism. Roughly, the argument is as follows, if zombies are conceivable, then physicalism must be false.


automatically think of someone morally omniscient. For instance, Brink gives the following account of a person omniscient about all natural facts:

He could presumably list all the obligations people would have in different sorts of circumstances and could then formulate extremely complicated moral rules specifying all of these obligations. Then one set of moral rules would apply to everyone, although, of course, they would tell people in different circumstances to do quite different things. This set of rules would state (at least a large part of) the one single true morality. We can even say that these realist claims amount to a form of moral absolutism; but this form of absolutism is no embarrassment to the moral realist.²

A naturalist like Brink thinks that complete physical and mental information about the world already includes moral information. This is why he thinks that if someone like Jeeves had enough time he could write not only a complete physical and psychological description of the world but also an account of the one single true morality. Brink and other like minded naturalists think that Jeeves cannot fail to know whether something is good. They also think that Jeeves's knowledge of all natural facts entails not only that he will make moral judgments but also that these judgments would be justified and true by virtue of what he knows about the natural world. As they see it, Jeeves's comprehensive knowledge of natural facts would enable him to see which natural properties moral terms and concepts pick out. Since Jeeves also knows all the natural features of every practical and substantive moral matter he will also know whether each of these is morally good.

So it seems that in the face of the moral naturalist's account of a person who knows all natural facts the first premise of the world book argument merely amounts to question begging. It does not help the antinaturalist's case that she could launch a begging the question charge of


Brink says that "if the moral realist were omniscient and had sufficient time [ibid.]" he would also be morally omniscient. But perhaps a more accurate statement of the case is as follows: if moral naturalism of the sort that Brink endorses is true, then a person who knows all natural facts would always know whether something is morally good.
her own against someone like Brink. In the same way that the naturalist could say that the proponent of the world book argument would only be able to say that fallible omniscience flouts the naturalist's expectations because such a scenario already implicitly assumes that moral naturalism is false, the antinaturalist could say that the only way that someone like Brink could say that someone who knows all natural facts would know the one single true morality is that he assumes that there is such a thing as the one single true morality. Charges of begging the question from both sides merely brings the discussion between the naturalist and the antinaturalist to a standstill. Unless one of the participants in the debate thinks of a better argument to prove their position or demolish their opponent's, the discussion cannot move forward. But perhaps it is possible to find something about which both the naturalist and the antinaturalist can agree so that an interesting discussion about the world book argument can proceed.

The naturalist and the antinaturalist can agree that moral theorists seek to account for for moral phenomena by making use of the information about moral phenomena afforded by the natural and social sciences. From here they can go on to say that if moral naturalism is true then a certain sort of explanation for moral phenomena would emerge, or is already beginning to emerge. What sort of explanation for moral phenomena could confirm the moral naturalist's claim? We could begin by saying what type of explanation would not do. Natural science's explanation for laughter in particular and comedy in general do not incline us to say that there are objective comedic
properties and that comedic properties are natural. That we respond to
certain states of affairs or verbal signals with spontaneous and instinctive
expressions of amusement do not imply that there is a set of natural
properties in the world that elicit amusement.

While it is true that human beings laugh as a matter of course and that
a full natural description of human and other animal life would be incomplete
without an account of laughter, what is funny is not a factual matter. This is
not because there is no such thing as funny or knowing what is funny, but
because knowing whether something is funny does not depend only on
knowing the existing dispositions of the relevant population and inferring
statistical norms from these. Knowing that something is funny implies having
the right sort of properly functioning neural mechanisms. Although we
ordinarily discuss and critically examine our comedic responses to things,
when we do this we are not performing an exercise in self-discovery or even
engaging in an inquiry into objectively occurring comedic properties. Rather,
we engage in a critical reflection aimed at helping us to make certain
decisions about which dispositions to keep and which ones to abandon.


By making systematic and long-term “observations of naturally occurring
laughter,” Provine and his research associates are able to describe the
acoustics of laughter, where and when we laugh, how we laugh and so on.
Their data consists in a number of laugh-logs in which people take note of all
the times in which they laugh throughout a given period and compiled
observations from many hours of eavesdropping on the conversations of
laughing people in public places. And it shows that laughter plays an
important role in social bonding, solidifying friendships and pulling people into
the fold, together with its converse: ostracizing outsiders or enemies. Here
are some of Provine’s more specific findings: (a) females in general are
laughers while males are laugh-getters; (b) most laughter is not a response to
jokes or other formal attempts at humor (e.g. one of the most typical prelaugh
comments is ‘I’ll see you guys later’). These findings and scientists’ attempt
to make sense of them do not incline us to believe that there are comedic
properties in the same way that there are properties like brittleness or
malleability.
After considerable debate and deliberation we may decide as individuals or as a community to alter our dispositions to laugh at something in this or that way. But we may not find a way to justify this decision definitively. Being funny is a natural property in some sense, but not the sense which the naturalist needs to defend her position. So a functionalist-natural account of what’s funny does not imply that the property of being funny is objective in a philosophically interesting way. If being morally good is like being funny, then moral properties must not be like uncontroversially objective properties like redness and brittleness.

To see this more clearly, compare our tendency to find things funny to our tendency to express ourselves using certain idioms or to think that certain turns of phrase are better than others. The capacity for language is something that virtually all humans share. We all have a knack for using certain sounds, written symbols, or even bodily movements to refer to things and to express ideas. Experts from different fields tell us that deep structures in the human brain are implicated in this ability to use language. Steven Pinker expresses the now well-known idea as follows:

Language is not a cultural artifact that we learn the way we learn to tell time or how the federal government works. Instead, it is a distinct piece of the biological makeup of our brains. Language is a complex, specialized skill, which develops in the child spontaneously, without conscious effort or formal instruction, is deployed without awareness of its underlying

---

4 The strategy being used here is not a reductio of natural properties to comedic ones, or some other dispositional property whose distinct metaphysical existence is questionable, e.g. gastronomic properties. For an interesting discussion of the failure of an antirealist attempt at a reductio ad absurdum of moral properties by pointing out parallels between it and gastronomic value properties, see D. Loeb, “Gastronomic Realism--A Cautionary Tale,” *The Journal of Theoretical and Philosophical Psychology* 23 (2003): 30-49. The material point for my discussion is that it is possible to find a naturalistic dispositional explanation for comedic properties and moral properties that defines the fittingness of some comedic or moral response that is neutral between value and disvalue. If this is the case, then moral naturalism cannot locate the kind of morally normative distinction that any moral theory needs in the descriptive.
logic, is qualitatively the same in every individual, and is distinct from more general abilities to process information or behave intelligently. For these reasons some cognitive scientists have described language as a psychological faculty, a mental organ, a neural system, and a computational module. But I prefer the admittedly quaint term “instinct.” It conveys the idea that people know how to talk in more or less the sense that spiders know how to spin webs. Web-spinning was not invented by some unsung spider genius and does not depend on having had the right education or on having an aptitude for architecture or the construction trades. Rather, spiders spin webs because they have spider brains, which give them the urge to spin and the competence to succeed.5

The language we speak then is a byproduct of a universal aptitude that past speakers in our native community have shaped and which present linguistic practices develop further. In his later writings, Wittgenstein expresses a similar idea as follows: “Our language may be seen as an ancient city: a maze of little streets and squares, of old and new houses, and of houses with additions from various periods.”6 It may be true then that some idioms are more apt for expressing this or that idea and that using certain phrases for certain purposes makes for better prose. But the explanation for the rightness of using such idioms and such turns of phrase is not the sort that the naturalist needs to support her position. It is only right to use such and such a word in the same way that it is right for some spider to spin its web a certain way because it is a member of a certain species and it exists at such and such an environment at such and such a time. Spiders evolve over time so that they begin to spin differently. The way people speak a language similarly changes over time through a combination of deliberate effort and


The similarities between the idea expressed in the quote from Pinker and Wittgenstein’s striking comparison between language and an ancient city may be incidental. It is merely being pointed out here that an idea that resonates with some remarks from Wittgenstein’s later writings fits well with the defense of WB1 being discussed in this chapter.
accidental alteration. And it may not be possible to find a way to systematically integrate the different ways that language is used and to find a single homogenous account of what counts as right usage.

The analogous explanation for morality that flouts the naturalist’s expectation is best expressed in the following suggestion:

Even if the human moral capacity evolved out of primate group life, this should not be taken to mean that our genes prescribe specific moral solutions. Moral rules are not etched in the genome. ...

We are not born with any specific moral norms in mind, but with a learning agenda that tells us which information to imbibe. This allows us to figure out, understand, and eventually internalize the moral fabric of our native society. Because a similar learning agenda underlies language acquisition, I see parallels between the biological foundation of morality and language. In the same way that a child is not born with any particular language, but with the ability to learn any language, we are born to absorb moral rules and weigh moral options, making for a thoroughly flexible system that nevertheless revolves around the same two H’s [Helping and (not) Hurting] and the same basic loyalties it always has.7

What de Waal says about morality bears a striking resemblance to what Pinker says about language. A shared psychological capacity or neural mechanism allows us to learn how to make moral judgments in the same way that a shared instinct allows us to acquire a language.8 In the same way that we can barely subsist without knowing some form of language, we barely survive without making moral judgments or subscribing to some form of moral code. But this does not necessarily imply that there is some moral code or set of moral judgments that is right for everyone. There may be too much variety in the kind of moral judgments that we make, in what counts as a fitting judgment of this sort and in what contingencies alter the fittingness of

7 De Waal, Primates and Philosophers p. 166-7.

8 There is a wide range of languages to learn. And language must be understood in its most capacious sense, e.g. including sign language. Also, what language or languages are learnt inevitably exerts an influence over how an individual sees the world. For a striking example see Oliver Sacks, Seeing Voices (New York: Vintage, 1989). These considerations make it even more difficult to find a single coherent account of what counts as the right way to use a word or a phrase.
making such and such moral judgments. So much so that it is not possible to find a synthesis of our moral practices that the moral naturalist is hoping to find.

De Waal’s hypothesis about morality is by no means an established empirical fact. Much is still needed to be done before we can even begin to see that it is true. But it is certainly possible that de Waal is on to something. The relevant point here is that both the naturalist and the antinaturalist can agree that if de Waal’s hypothesis about morality is right, then moral naturalism is in trouble. Here is the common ground that is needed for preventing the stalemate mentioned above.

The scientific explanation for seeing color implies that there is such a thing as a property possessed by some objects to produce certain sensations in the visual apparatus of certain animals, and that there is an objective property, namely the surface reflectance properties of objects, that explains color vision. That virtually all of us respond in a certain way to certain objects that produce certain sensations in our optical faculties demonstrates that color is indeed an objective property. Full information about how we talk about color and how we judge whether an object is this or that color would yield a single coherent account of color vision that could allow someone who has cognitive access to such information to know whether some object, say a ripe tomato, is this or that color. Such a fully informed individual may not know what it is like to see red because she has not actually seen a red object, but she would know that a ripe tomato is red. This is the type of scientific explanation for morality that the naturalist needs to vindicate her theory. If an empirically verifiable theory about morality turns out to be vindicated then fallible omniscience would turn out to be a kind of fantasy. Jeeves would be like flying pigs or a moon made of cheese, logically conceivable and
interesting but irrelevant. But until the relevant kind of empirical evidence becomes available, the competing naturalist and antinaturalist accounts of the same moral phenomena stand on the same level as plausible hypotheses.

Now that a common ground between the naturalist and the antinaturalist has been established, it is possible to discuss the world book argument. In the next section an explanation for Jeeve's moral ignorance that comes from the relevant type moral disagreement will be explored further.

Moral Disagreement and Knowledge of all Natural Facts

Many philosophers argue for moral non-cognitivism or moral antirealism, positions which are incompatible with naturalism, from moral disagreement. Some argue that the lack of convergence of opinion on moral matters implies that moral beliefs are not formed in the way that beliefs associated with sense perception are formed. Others say that widely diverging beliefs about many moral issues is a good basis for saying that moral terms are not referential in function.

---

9 The list of philosophers include C.L. Stevenson, R.M. Hare, J.L. Mackie, and Crispin Wright.

10 Mackie, *Ethics: Inventing Right and Wrong*, pp. 36-8. Mackie famously takes the wide variety of moral beliefs, and its implication that disagreements that feature these would be irresolvable, as a basis for saying that moral beliefs reflect a way of life rather than a perception of some part of the world. Mackie takes what he calls the relativity of moral beliefs as a basis for saying that moral terms are not referential in function.
moral terms have a non-referential function. Still others think that the rational intractability of disagreement on moral matters demonstrate that no moral belief is justified and that moral knowledge is therefore epistemically inaccessible. These different versions of the argument from disagreement take the semantic or epistemic implications of persistent disagreement on moral matters as a basis for saying that there are no moral properties, moral terms do not refer to anything objective, or moral knowledge is impossible to attain. Here it will be shown that moral disagreement of a certain sort can explain Jeeves's failure to know whether something is morally good in a way that is detrimental to moral naturalism. The supervenience of the moral on the natural ensures that the application of moral terms supervene on the distribution of certain natural properties. But since there is a significant range of variation on what these natural properties could be, a certain type of moral disagreement cannot be resolved by solely appealing to more natural facts.

11 Stevenson, Ethics and Language, pp. 59-71. Stevenson argues that moral disagreements are persistent and irresolvable because they involve not only disagreement in belief but also a disagreement in attitude. Stevenson takes what he calls the dual nature of ethical disagreement further, he argues that moral terms are chiefly emotive in function.

See also D. Loeb, “Moral Realism and the Argument from Disagreement,” Philosophical Studies 90 (1998): 281-303. In what he calls the semantic version of the argument from disagreement Loeb points out that “If people have widely differing beliefs about a number of moral questions, perhaps charity requires that we interpret them as referring to different properties, or using moral language in some entirely different, non-referring way (as non-cognitivists contend).” (p. 292)

12 Crispin Wright, Truth and Objectivity pp. 89-94. Wright is credited with putting forward a different version of the argument from moral disagreement. This version is usually called the argument from epistemic inaccessibility because it takes the persistence and intractability of moral disagreement as evidence for the inaccessibility of moral facts. Since it is absurd to postulate epistemically inaccessible facts, this argument is supposed to work as a reductio of moral realism. See also William Tollhurst, “The Argument from Moral Disagreement,” Ethics 97 (1987): 610-21. Tollhurst presents an earlier and more straightforward argument against moral realism based on the idea that one type of moral disagreement demonstrates that moral beliefs understood on the realist model are never justified.
The argument for the first premise of the world book argument that will be given here again makes an appeal to a specific type of moral disagreement, namely that type of moral disagreement that shows that the disputants are either both justified about their conflicting moral beliefs, or else none of them are. This type of moral disagreement is prefigured in Horgan and Timmons’s moral twin earth. Recall that in chapter 2 it was shown that the argument between earthlings and twin earthlings can be expressed in terrestrial terms because such a disagreement essentially depends on the conjoint presence of the following features: (1) the disputants use ‘good’ and right,’ or something similar, as their most general terms for moral appraisal and commendation; (2) the terms ‘good’ and ‘right,’ are used to reason about considerations that bear on well-being of individuals and entire communities or populations; (3) people are normally disposed to act in ways corresponding to what is ‘good’ and ‘right;’ (4) people take the goodness or rightness of options to be an overriding consideration in deciding what to choose; (5) the disputants apply the terms ‘good’ and ‘right’ to actions, persons, and institutions; and (6) the disputants use moral terms in such a way that their application supervenes on the distribution of natural properties.
of what they are disagreeing about. In this list, (3) and (4) capture what philosophers usually refer to as the normativity and the action-guidingness of moral terms. Taken together with (6) it makes a certain sort of moral disagreement possible in numerous instances.

A moral disagreement in which the disputants agree on natural or descriptive facts but disagree on what ought to be done because their use of moral terms is tracked by different sets of natural properties can arise in discussions of many other moral matters. The disagreement between A and B about capital punishment is not an isolated incident; it did not occur only because capital punishment is a controversial moral matter. The relevant type of moral disagreement can crop up in other places too. For instance, when people disagree about whether someone known to them is a good man much depends not only on how they conceive of good men and women but also on how they conceive of ‘good.’ Similarly, when people disagree about whether someone did the right thing, their dispute hinges not only on how they conceive of the action that was done but also on how their concept of ‘good’ works. If the people who disagree about these mundane moral situations...

---


Van Roojen takes most items from this list from what Horgan and Timmons themselves stipulate about the disagreement between earthlings and twin earthlings, although (1) and (6) are not explicitly stated in “Troubles for New Wave Moral Semantics: the ‘Open Question Argument’ Revived.” But van Roojen also says that he “doubts that Horgan and Timmons have listed all of the features of moral terms which account for our ascribing the same meanings to any terms used in the same way (p. 172).” Whether or not the list is complete does not have a direct bearing on the argument that I present here. For as long the example for moral disagreement that I present has all the features enumerated in the list and is plausible, it should be enough for the defense of the first premise of the world book argument that I will present.

14 See pp. 28-31.
matters conceive of moral goodness differently, e.g. if one of them thinks that moral goodness involves a duty to act only in accordance with a maxim that she can at the same time will to be a universal law while the other thinks that moral goodness involves promoting the greatest happiness of the greatest number, then their disagreement will persist in the same way that the disagreement between A and B about capital punishment would. It is true that not many ordinary folk can give an adequate account of the concept of moral goodness which is implicit in their use of ‘good’ and other moral terms, and that all but the few who have some knowledge of normative ethical theory would disagree about moral matters in this way. But it cannot be doubted that our ordinary conception of moral goodness allows for the kind of moral disagreement being described here. The material point is that in cases such as these it is hard to say which, if any, of the disputant’s moral beliefs is justified.

The justification of each person’s moral belief is tied to her conception of moral goodness and her use of moral terms, which both partly determine whether her moral beliefs are justified. The moral belief of a Kantian who thinks that capital punishment is morally forbidden will be justified in terms of some specific deontological theory. But since the same applies to her utilitarian interlocutor, it is difficult to see how the debate between them can be settled. For the purpose of settling the moral disagreement in question it is necessary to show through some suitably neutral standpoint which moral belief is justified. If neither moral belief can be justified in this way, or if both moral beliefs can be justified by an appeal to some partly neutral standpoint that imports the essence of the normative theory to which one party subscribes, then the disagreement remains unresolved.
Roughly, the problem that moral disagreement poses for moral naturalism is this: if the reason Jeeves could fail to know whether something is morally good is that most, if not all, moral beliefs are never justified from a naturalist standpoint then moral naturalism must be untenable. If moral naturalism is true then it must be possible to successfully arbitrate between disputants like A and B because it’s done all the time in natural science. This is why naturalists generally express optimism on long-term convergence of opinion on moral matters. For instance, Brink thinks that since there is virtual convergence of opinion about some moral matters like slavery, there is reason to believe that we would eventually reach an objectively secured consensus about most moral matters, just like we do about natural matters. Some moral disagreements may still persist but these would be the result of one of the following: (a) applying antecedently shared moral principles in different circumstances, (b) some moral agents hold systematically mistaken moral beliefs, (c) some cases involve options that have equal moral weight or incommensurable moral values, and (d) ignorance about some nonmoral, i.e. natural, matter of fact.\textsuperscript{15} But if Brink is right, then it must be possible to say how the moral beliefs about which we agree are justified in contrast with the way that other moral beliefs are not. Obviously, the naturalist cannot say that moral beliefs about slavery, torture for fun, etc., are justified because we agree about them. This would not amount to much. We agree that ripe tomatoes, blood, sunsets, etc., are red because they are red. They are not red because we all agree that they are.

Naturalists like Brink tell us that we gain access to natural moral properties in the same way that we gain access to other natural properties. Through observation and experience we become aware that certain objects

\textsuperscript{15} Brink, Moral Realism pp. 203-9.
are a certain color and that people who do certain actions have particular effects on our well-being. For instance, Nicholas Sturgeon thinks that naturalists can make an “appeal to the apparent causal role of ethical properties in the natural order. ... [since] Most of us can identify occasions in which we think we have benefited by someone else’s goodness or been harmed by their moral faults.” Sturgeon also points out that this conception of moral properties fits well not only with common sense but also with a long tradition of philosophical thought. Making inferences about the moral quality of actions and persons based on what we observe about them is a route to reliable moral knowledge according to the naturalist. Reasoning about these inferences and making mutual adjustments between them and widely accepted moral principles allow us to arrive at a more sophisticated understanding of morality, which can be tested through confirmational holism. Ethical beliefs cannot be tested in isolation from other ethical and non-ethical beliefs. We find the same thing in natural science where hypotheses can only be tested because they are informed by some widely accepted theory. So perhaps the relevant type of moral disagreement can be resolved by using this strategy.

But the strategy which emerges from the moral epistemology of moral naturalists would only work if the relevant analogy between ethics and natural science can be proved. It cannot work as a promissory note to the effect that we could expect the relevant type of moral disagreement would be resolved in a way that favors moral naturalism. This means that the naturalist has to show that what has worked so well for natural science does in fact work for

---

16 Sturgeon, “Ethical Naturalism,” p. 100.

17 For an account of naturalist coherentism see Sturgeon, p. 105-6 and Brink pp. 122-33. For an account of confirmational holism see Brink pp. 135-8.
ethics also. In order to do this, she would have to show that a complex theory, if not a paradigm, can triumph over its rivals in ethics in the same way that it does in natural science. This is where the disagreement between well-meaning and well-informed moral agents comes in. In such a case the disputants are unable to go beyond showing that their incompatible moral beliefs are consistent with their shared knowledge of descriptive facts. In terms of the moral epistemology of naturalists, what such a disagreement shows is that it is possible that two incompatible ways to understand morality cannot be arbitrated by confirmational holism. The reason for the possible failure of this method of settling moral disagreement may have something to do with what kind of explanation can be found for moral phenomena.\(^\text{18}\)

In the last section it was shown that moral naturalists need to establish that a certain type of naturalistic explanation for morality exists. If the reason for agreement on some moral matters turns out to be similar to the reason for the agreement about when it is appropriate to laugh or to use a

\(^{18}\) The possible explanation put forward here is of course inconsistent with what moral naturalists expect to find. Sturgeon thinks that agreement in scientific debates is limited in the following way: widespread agreement is only reached among “well-trained, well-informed, competent inquirers (p. 108).” So only the analogous moral case should matter. He seems to think that it is obvious to anyone that if we only count moral disagreement among well-trained, well-informed and competent inquirers settled results can be expects. He gives the example of homosexuality. The type of moral disagreement being discussed here is similar to the one that Sturgeon seems to have in mind. The difference about what this type of disagreement implies seems to depend on the difference between Sturgeon’s example and mine, i.e. homosexuality vs. capital punishment. The standoff between naturalism and antinaturalism threatens to resurface at this point. However, it is important to note that all that the first premise of the world book argument needs is an acknowledgment that the antinaturalist hypothesis about the relevant type of moral disagreement is on the same footing as the naturalist’s. Until more empirical information on the matter becomes available and/or a better argument for or against one position is put forward, the truth of the matter is uncertain. In any case, for present purposes what matters more is whether the second premise of the world book argument can indeed show that fallible omniscience, taken together with the explanatory hypothesis given here, implies antinaturalism.
certain word in a certain context instead of another, then moral naturalism cannot handle the relevant type of moral disagreement. So if our moral judgments are systematic and patterned in ways that are similar to the patterned occurrence of laughter that Provine observed or the universal occurrence of the capacity for language among humans that Pinker described, then Jeeves could indeed fail to know whether something is morally good and the reason for it is detrimental to the moral naturalist’s position.

Tollhurst’s version of the argument from moral disagreement captures the problem for moral naturalists very nicely. In order to see this, consider the following:

1. Necessarily, if person P is justified in believing that \( j \), then it is not possible for there to be another person Q who is similar to P in every epistemic respect relevant to being justified in believing that \( j \) (or not-\( j \)) and who is justified in believing not-\( j \).
2. Necessarily, if a person P’s belief with regard to proposition \( j \) is justified, then it is not possible for there to be a person Q who is similar to P in every epistemic respect relevant to being justified in believing that \( j \) (or not-\( j \)) and whose belief with regard to \( j \) is not justified.
3. Necessarily, if a person P’s doxastic attitude toward an objective moral proposition \( j \) does not result from a failure to reason correctly or any other (nonmoral) mental defect, then it is not possible for there to be a person Q
   a. whose doxastic attitude toward \( j \) does not result from a failure to reason correctly or any other (nonmoral) mental defect
   b. who holds the same nonmoral beliefs as P and is equally justified in holding them, and
   c. who is not similar to P in every epistemic respect relevant to being justified in believing that \( j \) (or not-\( j \)).
4. Therefore, if a person P is justified in believing that \( j \), where \( j \) is an objective moral proposition, then it is not possible that there be a person Q who meets conditions 3a and 3b and who believes that not-\( j \).
5. For every person P and for any objective moral proposition \( j \) that P believes, it is not possible that there be another person Q who believes that not-\( j \) and who meets conditions 3a and 3b.
6. Therefore, no objective moral beliefs are justified.\(^\text{19}\)

Tollhurst thinks that moral realists, who include naturalists, would be inclined to reject the third step of this argument. And he concedes that this premise is

too controversial to be taken as a definitive basis for rejecting moral realism. But Tollhurst points out that someone who rejects the third premise of this argument from moral disagreement incurs a significant theoretical liability since she must posit a moral sense. But for present purposes it is not necessary to delve into these aspects of Tollhurst’s argument. What matters is to that the argument relies on the kind of moral disagreement that is being discussed here and that Tollhurst’s argument from moral disagreement ties up nicely with point about the difficulty of finding some naturalistic way to establish that the moral judgments of one of the disputants in the relevant type of disagreement is justified.

Ultimately finding a way to show that some set of moral judgments are justified is like anticipating the possibility of massive convergence on moral matters. There are reasons for optimism and for pessimism about such a possibility. And we might only know who is right if more philosophically informed empirical research on moral disagreement becomes available. In the meantime it has been shown here that Jeeves as he has been described here is logically conceivable and that there are good philosophical reasons for taking such a conceivable possibility seriously. In the next chapter it will be explored further why this poses a problem for moral naturalism.

---

20 Even moral naturalists are divided on this matter. Nicholas Sturgeon is optimistic about the prospect of massive convergence on moral matters while Richard Brandt expresses doubt about the possibility of agreement about moral matters even under ideal conditions.


Richard B. Brandt, Facts, Values and Morality, p. 60.

Chapter 5
Sizing up the world book argument:

does fallible omniscience really imply that moral naturalism is false?

In his lecture Wittgenstein claims that since someone who knows all and only natural facts would still be morally ignorant, moral properties must either be supernatural or nonexistent. In the previous chapters this conclusion has been expressed in terms of the denial of the truth of moral naturalism. The premise of the world book argument that warrants this conclusion can be stated as follows: if someone who knows all natural facts does not know whether something is morally good, then the moral goodness of things must not be a natural matter. Wittgenstein does not defend this premise of his antinaturalist argument at length. He also seems to think that fallible omniscience demonstrates not only that moral properties are not natural but also that there are no moral properties:

If for instance in our world-book we read the description of a murder with all its details physical and psychological, the mere description of these facts will contain nothing which we could call an ethical proposition. The murder will be on exactly the same level as any other event, for instance the falling of a stone. Certainly the reading of this description might cause us pain or rage or any other emotion, or we might read about the pain or rage caused by this murder in other people when they hear of it, but there will simply be facts, facts, and facts but no Ethics.¹

Rather than state a simple moral antirealist conclusion as a natural consequence of his rejection of the belief that moral properties stand on the same level as natural ones, he says that if there are any ethical facts at all, these must be supernatural. Notwithstanding his preference for a position beset with serious problems,² Wittgenstein concludes that there are no

¹ LE pp. 6-7
² See pp. 41-42.
natural moral properties, and no natural moral facts. He takes the world book argument’s denial of the statement that moral properties are natural to be a reductio of moral naturalism.

The success of the natural sciences in establishing matters of fact, with its concomitant demand for parsimony, led to the wide acceptance of the platitude that if some putative fact or property has no cut and dried correlation with a definite set of natural facts or properties, then such putative facts or properties must not exist at all. Wittgenstein seems to think that since there is a discrepancy between a complete natural description of an ordinary, noncontroversial moral matter like stabbing an innocent person to death for some trivial, self-serving reason and the intuitive commonsense certitude that we have about the moral status of such an act, there must be something that stands between establishing a definitive correlation between this commonsense moral certitude and the natural facts and properties that could be enumerated in order to fully describe it. In other words, moral properties must not be natural if it is possible to fully describe even those morally charged situations about which we are certain, for instance murder, slavery and torture for fun, without necessarily coming to the conclusion that these things are morally wrong. In this chapter the strength of this line of reasoning will be examined more closely. One particular objection to the second premise of the world book argument will be considered. Roughly, the objection goes something like this: it is possible to account for the coherence of fallible omniscience and at the same time defend moral naturalism by saying that the reason Jeeves does not know whether something is morally good is that knowing all natural facts is compatible with failing to have some particular practical ability.
Natural properties versus moral properties

Before going on to a discussion of the objection just mentioned it is important to look more closely into the abstract principle behind the crucial premise of the world book argument. The second premise depends on the following idea: if someone can know all F facts without knowing whether X, which can be exhaustively described by enumerating F facts, is G, then ‘X is G’ must not be an F-matter. If Jeeves cannot infer from what he already knows that X is G, then the G-ness of X (if it really is the case) must not be an F-matter. The veracity of this simple abstract principle can be initially demonstrated by showing its relevance to delineating the different branches of natural science. If it is possible to fully describe some natural phenomenon by exhaustively enumerating its underlying chemistry and fundamental physics without necessarily arriving at the conclusion that such a phenomenon is a characteristic behavior of certain life forms, then the fact that a certain species behaves in this characteristic way must not be a chemical or physical matter.

For example, since there is no chemical or physical means to account for macrophysical entities such as individual organisms, species, genes, and so on, neither physics nor chemistry can allow us to see that members of species N synthesize the protein that they ingest in such and such a way. Even if it is true that the metabolism of all the members of species N works in a unique physical and chemical manner, a complete physical and chemical account of the matter could not afford us the kind of understanding that a biological account could. A complete description of species N in physical and chemical terms would not include the statement that members of this species behave in such and such a way in such and such circumstances. This is precisely because the characteristic behavior of species and the inner
workings of different kinds of metabolism is not a physical or chemical matter.\(^3\)

The example that has just been given can be replicated in terms of the difference between biology and psychology. A complete physical, chemical and even biological account of the stability of emotional responses in adults of a certain type cannot adequately account for what psychologists call a high aptitude for emotion-focused coping or emotional intelligence. However, the finer details of these examples from the different branches of natural science, which could perhaps be extended to the social sciences, need not bother us here. The material point is that even when one domain can be exhaustively described in terms of another, a definitive correlation between such a description of facts and properties in the domain with more complex macrophysical entities cannot be straightforwardly established. Of course these considerations do not prove that the abstract principle behind the premise establishes an insurmountable difference between moral properties and natural ones. That the different branches of the natural and social sciences are all natural in a broad but meaningful sense proves that merely establishing that morality can be delineated from the different natural and social sciences would not get the world book argument very far. What is more, all the different branches of the natural and social sciences are reducible to natural or physical terms. This is why it was mentioned above

\(^3\) This is not to say that biology is not reducible to chemistry, which is in turn reducible to physics. The possibility of establishing borders between the different branches of the natural sciences does not preclude the reduction of everything natural in terms of the physical. That such a reduction is possible only goes to show that the delineation between chemistry and biology is more for the purpose of taxonomy, and was perhaps devised for a better academic division of labor. The point of discussing the different branches of natural science in terms of increasing complexity rather than in reductive terms is that the abstract principle behind WB2 can be used to straightforwardly establish differences between related domains.
that these examples merely give the abstract principle behind the second premise of the world book argument initial plausibility.

A more philosophically relevant application of the abstract principle that stands behind the second premise of the world book argument is one that we can see in the analogous premise of the knowledge argument. This premise states that if someone who knows all physical facts about color vision does not know what it is like to see red, then physicalism about conscious experience must be false. Since Mary knows all the physical facts that can be enumerated to exhaustively describe the experience of seeing red and still fail to know what this experience is like, then such an experience must not be a physical matter. In Mary’s case the intuitive appeal of the supposition that she would not know what it is like to see red before she encounters a red object for the first time gives an additional compelling force to this conclusion. However, the material point about someone like Mary is not the intuitive appeal of her ignorance about what it is like to see red before she is released from her black and white room. It is the coherence of Mary’s situation that implies that there must be an insurmountable difference between a complete physical description of what it is like to see red on the one hand and the actual experience of seeing a ripe tomato on the other. That someone who knows all natural facts would fail to know what it is like to see red means that such knowledge must exceed knowledge of natural facts in some way. Unlike the previous examples from the different branches of natural science, this application of the principle behind the second premise of the world book argument hints at a more promising way of establishing a clear cut distinction between what is moral and what is natural. Only such a clear cut distinction could warrant the antinaturalist conclusion of the world book argument.
But how can an analogous clear cut distinction between morality and nature be established? This is where a comparison between Jeeves and Mary that focuses on the similarity between them comes in. Like Mary, Jeeves knows all the physical and mental facts that can be enumerated to exhaustively describe every morally good or bad thing. Also like Mary, he is ignorant about something which is closely related to what he already knows. So if it is possible for him to know all that and still fail to know whether some particular thing is morally good, then the moral goodness of that thing must not be entailed by knowledge of the natural facts that constitute it. Jeeves’s case may not be as intuitive as Mary’s; however, it cannot be denied that his situation is at least as coherent as hers. As mentioned in the previous chapter, the idea that someone who knows all natural facts is morally ignorant involves no contradiction. So if one accepts that Mary’s ignorance about what it is like to see red establishes an insurmountable distinction between this experience and physical facts about it, then one must also admit that fallible omniscience is a logical and coherent idea that provides a good basis for inferring that moral properties are not natural. That is, if one admits that the Jeeves’s situation is relevantly similar to Mary’s then one would have to conclude that the world book argument gives us good reason to think that there is an important difference between moral properties and natural ones.

An Ability Hypothesis About Jeeves

One standard reply to the knowledge argument is an ability hypothesis about Mary. According to those who put forward this hypothesis, the reason she “learns” something new when she is released from her black and white room is that she gains a new ability upon her first encounter with a colored object. David Lewis and Laurence Nemirow think that upon seeing a colored
object for the first time Mary could gain the ability to know how to imagine having the experience of seeing a colored object,\(^4\) or to remember, imagine and recognize colored objects.\(^5\) Both Lewis and Nemirow defend physicalism against the knowledge argument and at the same time account for the coherence of the possibility that Mary will learn something new when she finally sees a colored object for the first time by showing that what she learns can be construed as an ability that is already latent in all the physical information that she already possesses.

---

\(^4\) Nemirow discusses this objection against the knowledge argument in two of his articles:

“But some modes of understanding consist, not in the grasping of facts, but in the acquisition of abilities--for example, understanding a language, or understanding a rule (in Wittgenstein’s sense). As for understanding an experience, we may construe that as an ability to place oneself, at will, in a state representative of the experience. I understand the experience of seeing red if I can visualize red.

Now it is perfectly clear why there must be a special connection between the ability to place oneself in a state representative of a given experience and the point of view of experiencer: exercising the ability just is what we call ‘adopting the point of view of experiencer.’” [Review of Thomas Nagel’s Mortal Questions, Philosophical Review 89 (1980), p. 475.]

“Successfully visualizing a color, for example, engenders the ability to compare the color to other colors. So visualizing a color permits us to draw conclusions (or reason propositionally) about other colors as if we were seeing the imagined color. (I might conclude that the color I am imagining is a deeper shade of purple than the color I am witnessing.) Imagining seeing a color thus functionally represents seeing the color in our propositional reasoning about colors.” [“Physicalism and the Cognitive Role of Acquaintance,” Mind and Cognition: A Reader, Ed. William G. Lycan (Cambridge: Basil Blackwell, 1990), pp. 495-6.]

\(^5\) Lewis talks about an objection to the knowledge argument similar to Nemirow. Lewis’s version is expressed as follows:

“If you have a new experience, you gain abilities to remember and to imagine. ... By remembering how it once was, you can afterward imagine such an experience. Indeed, even if you eventually forget the occasion itself, you will very likely retain your ability to imagine such an experience.

The ability hypothesis about Mary is one form of a general response to the knowledge argument based on the idea that she does not gain any new propositional knowledge upon her release. This no-new-propositional-knowledge response explains away Mary’s prerelease ignorance about what it is like to see red by pointing out the limits of having cognitive access to propositional knowledge only. Knowledge that does not necessarily translate to knowledge how, even though the latter is already contained within the former.

Similarly, those who want to accommodate the coherence of the moral fallibility of someone who knows all the natural facts and at the same time defend moral naturalism against the world book argument might propose an ability hypothesis about Jeeves. They could say that the only reason why Jeeves could know what he does and still fail to know that murder or torture for fun is morally wrong is that he does not have some practical ability which is already latent in what he knows. An ability hypothesis about Jeeves’s moral ignorance amounts to saying that the absence of an account of moral properties in his world book can be explained away by the claim that he lacks some kind of skill or practical ability. A moral naturalist who wants to defend this hypothesis would say that Jeeves’s moral fallibility implies that the ability to know whether something is morally good does not depend on propositional knowledge about natural facts only. Rather, it depends on some kind of ability; for instance, the ability to know which natural facts are morally relevant. So the first premise of the world book argument does not imply that there are no moral properties. It merely implies that knowing whether something is morally good does not solely depend on having cognitive access to the relevant propositional knowledge, or even all such knowledge.
The proponent of the ability hypothesis has it that Jeeves is morally fallible only because he does not have certain skills or abilities. But what kind of ability could Jeeves lack? It's possible that he cannot imagine himself to be confronted with scenarios in which people typically make moral judgments which they take to be certain, e.g. murder, torture for fun, slavery, etc. His inability to imagine himself to be confronted with these scenarios could make him fail to see which natural features of such scenarios are relevant to its moral badness. Without an imaginative capacity that we all take for granted, Jeeves could fail to distinguish between morally relevant natural information and irrelevant information. Perhaps an integral part of the imaginative capacity that Jeeves lacks is the ability to take the different perspectives of those involved in murder, torture for fun, or slavery. Since Jeeves knows all natural facts he may not know what it is like to know things from a limited perspective and how to use information from such perspective taking to make moral judgments. So unless Jeeves knows what we go through when we imagine youths dousing a cat with gasoline and setting it on fire he cannot know whether something is morally good or bad. Or again, unless Jeeves knows what we go through when we see someone torture a cat or a human being for no apparent reason translates to knowing that such acts are morally wrong is not clear. It is not obvious, for instance, that having firsthand knowledge of outrage and disapproval over what the youths did to the cat would necessarily translate to having a desire that such acts not be perpetrated. Even if Jeeves does not have firsthand knowledge of outrage and disapproval of torture, he has secondhand access to such information. Jeeves even knows that virtually all of us are outraged by and disapprove of torture. If all this information is not sufficient for knowledge that torture is wrong, or what comes to the same thing: the judgment that torture is not morally good, then moral knowledge must not be solely a function of possessing such natural information. But for the sake of argument and for the purposes of developing an ability hypothesis about Jeeves I am supposing here that the judgment that what the youths did to the cat is morally wrong could be secured by the moral naturalist if she also provides a suitable account of the connection between knowing what we experience when we see such an atrocity perpetrated and knowing that such an act is morally wrong.

---

6 How knowing what we go through when we see someone torture a cat or a human being for no apparent reason translates to knowing that such acts are morally wrong is not clear. It is not obvious, for instance, that having firsthand knowledge of outrage and disapproval over what the youths did to the cat would necessarily translate to having a desire that such acts not be perpetrated. Even if Jeeves does not have firsthand knowledge of outrage and disapproval of torture, he has secondhand access to such information. Jeeves even knows that virtually all of us are outraged by and disapprove of torture. If all this information is not sufficient for knowledge that torture is wrong, or what comes to the same thing: the judgment that torture is not morally good, then moral knowledge must not be solely a function of possessing such natural information. But for the sake of argument and for the purposes of developing an ability hypothesis about Jeeves I am supposing here that the judgment that what the youths did to the cat is morally wrong could be secured by the moral naturalist if she also provides a suitable account of the connection between knowing what we experience when we see such an atrocity perpetrated and knowing that such an act is morally wrong.
has the ability to imagine what it is like to be both the killer and the killed he may not know which attendant natural features are relevant to the wrongness of murder.

It is also possible that Jeeves does not have the ability to cognitively process morally relevant affect in the way that we do. Without this ability Jeeves would fail to recognize which among the many things that he knows fall under the rubric of morality. Morally relevant affect allows us to see which actions or states of affairs fall within the domain of morality. Guilt, envy or desire alerts us to the possible moral significance of the thing which elicits such affective responses. And the ensuing cognitive processing of these responses are implicated in knowing whether something is morally good. If Jeeves does not have both the mechanism for having affective responses and the ability to cognitively process these responses, then it is no wonder that he does not know whether the willful killing of an innocent for trivial, self-serving reasons is morally wrong. The ability to cognitively process morally relevant affect could allow Jeeves to situate himself within the moral domain so that he would see murder for what it is. But since he does not have the requisite ability he fails to recognize not only that murder is wrong but also that there is a domain in which murder, torture for fun and slavery have a distinctive moral status.

Taken together, the two abilities that Jeeves putatively lacks present a formidable challenge to the proponent of the world book argument.

Reply to the Ability Hypothesis About Jeeves

A first line of defense against the ability hypothesis about Jeeves consists in adopting the strategy employed by proponents of the knowledge argument who say that the proposed ability that Mary putatively lacks is
neither necessary nor sufficient for knowing what it is like to see red.\textsuperscript{7} If the
ability to know how to imagine having the experience of seeing a colored
object and the ability to remember, imagine and recognize colored objects are
indeed neither necessary nor sufficient for knowing what it is like to see red,
then it is irrelevant to say that Mary does not have these abilities. These
abilities can only explain away the coherence of the idea that Mary will learn
something new when she sees a colored object for the first time if they are
necessary and sufficient for knowing what it is like to see a colored object. If
the putative abilities are not related to knowing phenomenally red properties,
then they are not relevant to the knowledge argument at all.

A proponent of the world book argument can adopt this strategy for
Jeeves. She can also claim that the abilities that he putatively lacks are
neither necessary nor sufficient for knowing whether something is morally
good. For instance, someone with limited imagination but extensive
experience of morally relevant cases would know whether something is
morally good, especially if the thing in question is something as obvious as
murder or torture for fun. Even if she has not encountered that particular
case before, what she already knows would most likely allow her to see that
the case before her resembles a number of cases about which she is certain
in different ways. From this she can easily form a judgment about not just the
novel case before her, but to most new cases that she might encounter.
Jeeves is an even better informed position compared to such a hypothetical
moral agent. He knows not only all the morally relevant cases but all moral
cases that there are. So even if Jeeves does not have the ability to imagine
himself to be confronted with different types of morally charged scenarios, or

\textsuperscript{7} For this response to the ability hypothesis about Mary see Earl Conee,
138-9.
even the ability to take the perspective of those involved in them, he must always know whether something is morally good. That is, if there is any fact to know about a morally charged scenario, Jeeves ought to know them as a matter of necessity.

The other ability that Jeeves putatively lacks is the ability to cognitively process morally relevant affect. Presumably such an ability would only be helpful to the moral naturalist if the cognitive processing of affect somehow tracks objectively existing moral properties, and if it were really the case that this proposed ability functions in this way then Jeeves would know it. He is not limited to our epistemic access to these putative objectively existing moral properties. If the cognitive processing of affect is indeed our unique mode of access to objectively existing moral properties, then there must be some other way to access these properties. And if there really are objectively existing moral properties then Jeeves is in an even better position to apprehend them. He would not be prone to make mistakes about processing affect, which are a potential source of distorting bias, because he can access the putative moral properties through what he knows about our mental capacities and the objects which these apprehend. Neither is Jeeves limited to the perspective of an individual or a community’s cognitive processing of morally relevant affect. He knows about all the different instantiations of such processing, so if there really are objectively existing moral properties that underlie these instantiations of cognitive processing of morally relevant affect, then Jeeves must be able to perceive them.

In the end, both of the abilities that Jeeves putatively lacks could only become helpful to the moral naturalist who wants to block the second premise of the world book argument if it satisfies the following conditions: (a) the abilities are necessary for knowing whether something is morally good; and
(b) these abilities are established as our unique mode of access to objectively existing moral properties. Fulfilling (a) demands considerable philosophical work; fulfilling (b) appears to be an empirical matter. Satisfying both (a) and (b) is tantamount to proving the truth of moral naturalism. So it seems that the moral naturalist has to win the debate against her antinaturalist opponent first before this objection to the world book argument can be posed. What is more important, even if both conditions were fulfilled the ability hypothesis about Jeeves would only work if the moral naturalist also denied the coherence of the possible moral fallibility of the person who knows all natural facts. Recall that the appeal of the ability hypothesis about Mary consists in its power to explain away the coherence of the idea that she will learn something new when she is released from her black and white room while keeping physicalism intact. In the case of the ability hypothesis about Jeeves, it could only manage to keep moral naturalism intact if a separate objection to the first premise of the argument is also posed. So it’s beginning to look like posing an ability hypothesis about Jeeves is not a very effective way of blocking the second premise of the world book argument.

The best line of defense against an ability hypothesis about Jeeves is to combine the above-mentioned points with an account of an important contrast between philosophy of mind and ethics. Previously, only the similarities between the knowledge argument and the world book argument have been discussed. Although the strategy behind the two arguments is essentially the same, the difference between the domains in which they figure cause them to bring about widely diverging conclusions. In the philosophy of

---

8 The moral naturalist cannot simply assume that the cognitive processing of morally relevant affect is indeed our unique mode of access to objectively existing moral properties because this would presuppose that these properties exist. Merely assuming that there are such properties amounts to a *petitio*. 
mind the knowledge argument leads to dualism, which amounts to the claim that phenomenally red experiences are metaphysically discrete from the physical mechanisms from which they arise. But in ethics the world book argument brings about an antinaturalist conclusion, which amounts to a denial of the existence of natural moral facts or properties. In the philosophy of mind the reasoning is as follows: if Mary does not know about it before she is released, then it must have a distinct metaphysical existence from what she already knows; in ethics, the reasoning goes something like this: if Jeeves is ignorant about it, then there must be no such thing. This difference can be attributed to the difference between what Jeeves and Mary know and are ignorant about.

Knowing what it is like to see red is different from knowing whether something is morally good in that the former is an indubitable concomitant of visual perception whereas the latter is not an indubitable concomitant of knowing natural facts. Whereas the existence of phenomenally red experiences cannot be doubted, the existence of moral properties, together with the possibility of moral knowledge, is controversial. That is because ethics and the philosophy of mind involve distinct sets of assumptions. When it comes to conscious mental experience the point of departure is the unique quality of our subjective or introspective access to seeing colored objects, especially when it is contrasted with an objective and impersonal account of what is going on when we have this experience. Whether our unique phenomenal experiences supervene on knowledge of the relevant physical facts is what is up for debate. But when it comes to ethics it is not assumed at the outset that there are objectively existing moral properties and that it is possible to know about these. Whether there are such things as moral properties is up for debate. Instead, the point of departure in ethics is the
supervenience of the moral on the nonmoral. The crucial question is not whether Jeeves lacks abilities that we take be crucial to our knowing whether something is morally good, rather the question is whether the abilities that we need to participate in the practice of making moral judgments and the activity of theorizing about these are connected to natural facts in the right way.

Because of the difference between ethics and philosophy of mind, postulating similar hypotheses about what Jeeves and Mary do not know have very different metaphysical implications. The possibility that Mary will not know what it is like to see red casts doubt on the supervenience of the mental on the physical, but it does not cast doubt on the existence of phenomenally red experiences. If anything, the very existence of this unique experience is affirmed in the knowledge that it exceeds any complete physical account of it. In Mary’s case physicalists think that her access to all the information about all the physical base properties of seeing red would also afford her access to the supervening phenomenally red experiences. Dualists, on the other hand, think that despite the irrefutable connection between physical information about color experience and the unique subjective quality of this experience, the former is discontinuous with the latter in an important way. They claim that our tendency to think that Mary will not know what it is like to see red until she actually sees a red object under ideal viewing conditions with adequately functioning optical faculties, points to the metaphysical distinctness of phenomenal properties from physical ones. So if Mary learns something new when she is released from her black and white room this would mean that there is some aspect of the mental that exceeds the physical. Regardless of which position one takes, the existence of phenomenal properties associated with seeing red is assumed to be indisputable.
But in the moral case, if Jeeves would fail to know whether something is morally good despite being fully informed about the natural features of that thing, this does not imply that moral properties transcend natural ones. Since the existence of knowing whether something is morally good does not stand on the same level as the existence of phenomenally red experiences, failure to derive the former from knowledge of natural facts implies that there might be no such thing as moral properties at all. The reason Jeeves could fail to know whether something is morally good might be that there is no moral fact to know about at all. This is a possibility that moral supervenience allows for. Within morality the debate is not about whether moral supervenience is true but about how best to account for it. Moral realists believe that the moral supervenes on the descriptive because moral properties are descriptive properties. They think that the necessity of saying that two things with the same descriptive or natural features have the same moral worth shows that the latter stands on the same level as natural properties. Antirealists, on the other hand, think that the supervenience of the moral on the descriptive is a limit on what we find intelligible as a sincere utterance of a statement that features a moral predicate. For the antirealist, moral supervenience is a conceptual constraint on what kinds of prescriptions, expressions of attitudes, etc., are moral judgments. According to antirealists, that virtually all of us share the same noncognitive responses towards certain things, e.g. murder, torture for fun, etc., shows that a long history of repeated projection of our
attitudes over such acts have sedimented into established convention,\(^9\) not that natural moral properties exist.

However, it is necessary for both realists and antirealists to account for the fact that the supervenience of the moral on the nonmoral or descriptive is a conceptual necessity that is constitutive of morality. It is a platitude that ethical theorists must account for. Whatever their stance on the existence of moral properties or their position on what kind of property moral properties are, realists, antirealists, naturalists and antinaturalists alike must say why the moral supervenes on the descriptive. The merit of their position depends in a large part on the plausibility of the explanation that they provide and its coherence with the rest of the tenets of their position. One problem that confronts the moral naturalist is that her claim that the moral supervenes on the descriptive because moral properties are natural properties is yet to be fully explained. For instance, it is not clear what is meant by ‘because’ in this moral naturalist claim. The putative equivalence, or at least objective correlation, between moral properties and natural ones is based on yet to be found empirical evidence. For all we know such empirical evidence can be found. But in the meantime we already know that the moral supervenes on the descriptive and that someone like Jeeves is logical conceivable so such knowledge must not depend on the empirical evidence that the moral naturalist is still hoping to find.

Moreover, Jeeves must know about moral supervenience; he cannot know what he does without knowing that we take the supervenience of the

\(^9\) For instance, our shared deep evolutionary past may have shaped us in such a way that we approve of cooperative behavior towards conspecifics, especially our own kin, and that we disapprove of unprovoked murderous hostility and needless infliction of pain. It is true that this fact can be appropriated by moral naturalists and other moral realists for their own ends. But they would need to say more about the exact nature of the relation between moral properties and natural ones before they could do this.
moral on the descriptive to be constitutive of correct usage of moral terms and concepts. The naturalist contends that moral properties are natural and that the right way to account for moral supervenience is to say that moral properties supervene on natural ones in the same way that, say, biological properties supervene on physicochemical ones. But since the naturalist cannot say exactly how moral properties are related to descriptives ones then she cannot say either how Jeeves could, or even would necessarily, derive moral knowledge from what he already knows. This means that talking about a purported ability to derive moral knowledge from knowledge of natural facts, which Jeeves putatively lacks, would be problematic also. So the moral naturalist’s ability objection to the second premise of the world book argument comes to naught.

Simon Blackburn even goes so far as to say that since moral realism, which encompasses moral naturalism, cannot account for moral supervenience as well as antirealists could, there is at least one good reason
to reject moral realism.\(^\text{10}\) The reasoning behind Blackburn’s supervenience argument against moral realism is as follows: if it is true that through the methods of natural science we would eventually discover which natural properties are objectively correlated with moral properties, then moral supervenience appears to be mysterious, or else an inexplicable brute fact. In contrast, expressivism fits well with moral supervenience. According to the expressivist we project values onto the world; hence value comes before experience. The supervenience of the moral on the nonmoral or descriptive is a logical constraint upon our practice of projecting value, so just like the values that we spread onto the natural world moral supervenience precedes experience.

But it is not necessary to endorse Blackburn’s supervenience argument against moral realism, or to subscribe to expressivism, to show that the second premise of the world book argument is true. It is enough to

\(^{10}\) Since he is an expressivist, someone who believes that ethical statements do not express beliefs, Blackburn demonstrates that unlike the moral realist he can adequately account for the a priori supervenience of the moral on the nonmoral. Roughly, his supervenience argument against moral realism is as follows:

\[
\begin{align*}
S1 & \quad \text{It is a priori that moral properties supervene on natural properties.} \\
S2 & \quad \text{If it is a priori that moral properties supervene on natural properties, then a good metaethical theory must explain this.} \\
S3 & \quad \text{Moral realism cannot adequately account for the supervenience of the moral on the natural.} \\
S4 & \quad \text{Expressivism can adequately account for the supervenience of the moral on the natural.} \\
S5 & \quad \text{So the a priori moral supervenience of moral properties on natural properties gives us a reason to accept expressivism and reject moral realism.}
\end{align*}
\]

recognize the significance of a fundamental presupposition about morality that makes Blackburn's supervenience argument possible. Moral supervenience is not only neutral between moral realism and moral antirealism, it takes precedence in the debate between the proponents of these two positions. Since Jeeves knows all natural facts he must also know about this feature of our practice of making moral judgments and theorizing about this practice. Since it is possible for him to know all natural facts including this fact about our practice of making moral judgments and still fail to know whether something is morally good, then moral naturalism is in trouble.

Knowledge of all the natural facts gives Jeeves the ability to have the kind of higher order knowledge of natural matters that physicists, chemists and biologists have about particles, elements, compounds and organisms. But if Jeeves cannot reliably develop moral expertise from what he already knows, then the ability to tell the difference between good and bad must not be analogous to having such higher order knowledge of natural matters. Presumably, Jeeves can formulate an intricate psychological account of our world based on what he knows about the fundamental natural facts that obtain. This proves not merely that the mental is part of the natural world but also that the mental is an objective domain in its own right. If as the first premise of the world book argument contends it is possible for Jeeves to know all that he does and still fail to know not only that the willful killing of the innocent is wrong but also that there is some moral domain about which there are laws that are as hard as any in the natural and social sciences, then there really must be no such thing as moral properties.
Despite what has already been said in support of the second premise of the world book argument, some might still think that it is possible to accommodate fallible omniscience within a sophisticated moral naturalist framework. Surely, they say, it is possible to admit that Jeeves would fail to know whether something is morally good without denying that moral properties are natural. In other words, surely it is possible to block the second premise of the world book argument and still take Jeeves seriously. An ability hypothesis about Jeeves might not be sufficient to block the relevant premise but surely some other hypothesis can get the job done. To put forward such a hypothesis, the naturalist must give an explanation for Jeeves’s failure to possess the kind of natural information that would enable him to know whether something is morally good despite his omniscience about natural facts. That is, she must explain why morally relevant natural information is inaccessible to Jeeves as such even though he knows all natural facts.

The naturalist has it that the moral worth of something is determined by its morally relevant natural features. But how natural features present themselves as morally relevant may not be a straightforward matter. Perhaps the best way to pursue this naturalist response to the world book argument is to talk about the distinctive presentation or appearance of morally relevant natural information. A naturalist may even hypothesize that knowing that something is morally wrong entails having some distinct relation to a natural
fact. In other words, there might be a unique way of conceptualizing or being related to natural information available to someone who knows whether something is morally good that is not available to Jeeves. Of course the naturalist would also have to explain why the sort of conceptualization available to common folk is unavailable to Jeeves. But if she can find such an explanation, then she could show that despite being interesting and to some extent plausible the world book argument does not really undermine moral naturalism. In what follows, such a response to the new antinaturalist argument will be examined in great detail.

The new guise hypothesis about Mary

One common response to the second premise of the knowledge argument closely matches what the naturalist needs to mount the requisite

\footnote{For a development of this response against both the knowledge argument and the world book argument see Michael Pelczar, “The Knowledge Argument, the Open Question Argument and the Moral Problem,” *Synthese* 171 (2009): 25-45. In this article an argument similar to WBA is discussed in terms of an ‘old relatum/new relation’ strategy. This strategy is related to the new guise hypothesis, which is going to be considered here, in the following way: whereas the proponent of ‘new guise’ claims that what prerelease Mary lacks is knowledge of all the different ways of conceptualizing ontological facts, the proponent of ‘new relation’ contends that what someone like Jeeves lacks before his moral ignorance is cured is a desiderative relation to some natural fact, e.g. a desire that murder, torture, etc. not be perpetrated. The proponent of new relation assumes that knowledge of some moral fact, if there be such a thing, is none other than a desiderative relation to some natural fact. By discussing the situation of someone who knows all natural facts, Carrie, Pelczar demonstrates that the second premise of the world book argument is false:}

"Carrie knows every fact ... is consistent with her failing to desire, for example, that genocide not be perpetrated. But if Carrie does not have this desire, then she does not know that it is wrong to perpetrate genocide. (This is so even if the fact that it is wrong to perpetrate genocide is some natural fact about genocide that Carrie knows.) So, the fact that Carrie knows every fact--including every moral fact, if there are moral facts--is consistent with her failing to know that it is wrong to perpetrate genocide. But if she does not know that it is wrong to perpetrate genocide, then she suffers from moral ignorance. ... Therefore ... even if all facts are natural facts, someone could know all the natural facts, yet suffer from moral ignorance (p. 34)."
objection against the analogous premise of the world book argument.² Like
the ability hypothesis, this objection to the second premise of the knowledge
argument concedes that Mary would learn something upon her release, but
unlike the ability hypothesis it also concedes that what Mary learns is
propositional knowledge. The crucial move for this objection is the denial that
the information that Mary gains when she encounters a colored object for the
first time is altogether new to her. This means that those who pose this
objection deny that the new information that Mary learns is something above
and beyond the physical facts about color vision that she already knows.
Since those who advance this response say that what Mary learns when she
finally goes out of her black and white room is a new guise of an old fact that
she already knew, this strategy is sometimes called the new guise hypothesis
about Mary. According to this hypothesis, what Mary learns when she sees a
ripe tomato for the first time merely consists in knowing about a new
presentation or a different packaging of something already known to her.
Roughly, the idea is that upon seeing a colored object for the first time Mary
will learn a new way to conceptualize an old physical fact because she will
finally learn firsthand what phenomenal redness is like.

² The relevant premises of the two arguments are as follows:

\[
\begin{align*}
K2 & \quad \text{If someone who knew all physical facts about color vision does not know what it is like to see red, then physicalism about conscious experience must be false.} \\
WB2 & \quad \text{If someone who knows all natural facts does not know whether something is morally good, then moral naturalism must be false.}
\end{align*}
\]
According to Horgan, one of the proponents of the new guise hypothesis, Jackson’s knowledge argument commits a fallacy because it equivocates on two different senses of physical information. Horgan has it that what Mary learns when she sees a ripe tomato for the first time can be expressed by a sentence that features an indexical term:

(R) “Seeing ripe tomatoes has this property;” where ‘this property’ is used to designate the property of phenomenal redness that is instantiated in her experience. Horgan adds that (R) “expresses new information because Mary has a new perspective on phenomenal redness: viz., the first-person ostensive perspective.”

Being sequestered in a black and white room merely prevents Mary from having access to a distinct perspective of a physical fact about color, namely the fact that a first person experience of seeing it has a distinct phenomenal property. Before her release Mary knows all about phenomenal redness in terms of the first person reports of people who have seen color and the corresponding brain states of these people. But since Mary herself has never seen a colored object she cannot point to that experience from the first person point of view. When she sees the sky for the first time she becomes more fully enlightened about what her informants are talking about when they see a colored object. But what Mary learns isn’t the type of new information

---

3 Proponents of the new guise hypothesis about Mary include Bigelow and Pargetter, Churchland, Loar, Lycan, Perry and many others. But for the purpose of adopting the new guise hypothesis for moral naturalism only Horgan and Tye will be discussed here. Horgan and Tye’s versions of the new guise hypothesis about Mary are the most relevant for Jeeves because their approach does not hinge on technicalities about phenomenal properties. As noted in the previous chapter, different assumptions about phenomenal properties and moral properties yield different conclusions about their existence. So strategies that depend on theoretical innovations on the character of phenomenal properties are mostly irrelevant to the naturalist versus antinaturalist debate in metaethics.


5 Ibid., p. 151
that is detrimental to physicalism about conscious experience because it corresponds to a fact already known to her from objective information about brain states and from secondhand reports from her informants.

Horgan points out that it is possible to individuate physical information in two ways: (a) in terms of the ontological objects that physical concepts correspond to; or (b) in terms of all the different physical concepts there are, even if some of these correspond to the same ontological object. So Mary can be said to be fully informed about the physical information on color vision in two ways. Either she knows all ontological objects in at least one guise or she knows all the different concepts that correspond to these objects. If all the knowledge argument shows is that Mary would learn something new when she finally gets released because the second premise of the argument is trading on the distinction between ontological objects and the concepts which correspond to these, then physicalism about conscious experience remains intact. All the physicalist needs to defend her position is to show that all mental concepts correspond to some physical object. That some mental concepts are only accessible through a distinct first person perspective does nothing to show that there are mental concepts that go beyond the physical. So Mary and others like her are not counterexamples to physicalism about conscious experience.

Michael Tye parses the new guise hypothesis about Mary in terms of an improvement on the ability hypothesis proposed by David Lewis. According to Tye, the information that Mary learns when she sees a colored

---

object for the first time is best captured by a disjunction of introspective knowing-that and an ability that can be summed up as follows:

“S knows what it is like to undergo experience E =df Either S is now undergoing E and S has indexical knowledge-that with respect to the phenomenal character of E obtained via current introspection or S has the Lewis abilities with respect to E.”

Tye improves on Lewis’ response to the knowledge argument by tying up the cluster of abilities associated with knowing what it is like to see red\(^8\) with the more basic ability to apply an indexical concept to the phenomenal character of an experience of phenomenal redness via introspection. What Tye calls the ‘more basic’ ability is inseparable from a direct acquaintance with phenomenal redness.\(^9\) According to Tye, even if prerelease Mary is fully aware that there is a distinct phenomenal concept ‘experience of red’ and that she can use such a concept when she is not talking about herself, she does not know what it is like to experience red until she actually does so. From this it follows that when she is still trapped in a black and white room Mary will not be able to say or think to herself, while actually seeing a red object, that this is what it is like to see red. Such a statement or thought remains inaccessible to her because its meaning or content is only accessible to someone who has seen or is seeing a red object.

Like Horgan, Tye counts what Mary learns when she sees a colored object for the first time as a new mode of presentation of an old fact.

\(^7\) Ibid.

\(^8\) Recall that the particular abilities proposed by those who pose the ability hypothesis about Mary include the ability to imagine, recognize, or remember color or color experiences.

\(^9\) In this respect, Tye’s version of the new guise hypothesis closely resembles what is sometimes called the acquaintance hypothesis about Mary. For a version of the latter that fits well with new guise, see Earl Conee, “Phenomenal Knowledge,” Australasian Journal of Philosophy 72 (1994): 139-41.
Similarly, Tye associates the new concept that Mary learns with indexical thought contents associated with seeing colored objects. Tye also denies that conceptual differences necessarily mirror worldly differences. According to Tye, “indexical thoughts and thought contents are partly individuated by the items picked out by the relevant indexical concepts and partly by concepts or modes of presentations themselves.”

This means that when Mary sees a ripe tomato for the first time and exclaims “So this is what it is like to see red!” the contents of her thought at that particular moment is the peculiar introspectible aspect of her experience. But at the same time Mary immediately sees the connection between what she learns and what she already knows. She will discover for herself the basis of the reports of her informants, and the subjective quality of being in the brain state of people who see red objects. If she is so inclined Mary herself will see that what she learns is already latent, although not fully realized, in what she knew in her black and white days. If this is right, then it could be true that, even though a whole world of color experience is unavailable to sequestered Mary, physicalism is still right about conscious experience. The naturalist could apply this old fact/new guise strategy to Jeeves. If successful, moral naturalism could remain intact despite fallible omniscience.

A new guise hypothesis about Jeeves

To pose a new guise hypothesis about Jeeves, the naturalist needs to identify something that ordinary folk do as a matter of course that could plausibly be inaccessible to Jeeves. What the naturalist needs to find is the equivalent of seeing ripe tomatoes, sunsets, blood, etc., for Mary, with respect to Jeeves. Unlike Mary most of us do not know, or even come close to

---

10 Tye, “Knowing What it is Like,” p. 228.
knowing, all physical facts about color vision. But at every moment we have access to the ostensive first person perspective on color. So there is distinct conceptualization of color that we know about that is unavailable to prerelease Mary. If the naturalist can find such an analogue for Jeeves, then it would be possible for her to talk about the distinct guise of morally relevant natural information and thereby block the second premise of the world book argument.

Jeeves, as he has been described thus far, is similar to sequestered Mary in the following way: he is disconnected from the way that moral agents experience and relate to the natural world. In the same way that Mary is removed from our ordinary access to color, Jeeves is removed from our ordinary access to the natural features of things that we judge morally. Jeeves’s knowledge of all natural facts implies that he does not have to rely on the methods that we employ to know the natural features things to know about what he is judging. For instance, he does not have to exert effort to find facts or to gather data about some particular case in the way that we do. Jeeves already knows all the natural facts that there are so he does not have to sift through information or gradually piece together the big picture in the way that we do. His omniscience also implies that he is not subject to the vicissitudes of misinformation in the way that we are. Someone who wants to propose a new guise hypothesis about Jeeves can use these similarities between him and Mary to explain why it is that he would fail to know whether something is morally wrong. Jeeves, as he has been described thus far, is not sequestered in the same way that Mary is. But being completely detached from the way that we experience the world is certainly consistent with fallible omniscience. This is all that a proponent of the new guise hypothesis about Jeeves needs to introduce an explanation for Jeeves’ failure
to know whether something is morally wrong that is consistent with moral naturalism.

Although knowledge of all natural facts does not imply detachment from our experience of the world, the latter is a plausible explanation for fallible omniscience that is consistent with moral naturalism. Jeeves is not apt to be mislead by facts because he does not know the ‘full story’ about something in the natural world. But perhaps for the very reason that he does not have to rely on a firsthand account of the natural features of things, he does not know what it is like to be confronted with some act, person or state of affairs in a context in which he has to decide what to do and what to think about what stands before him. Neither does Jeeves know what it is like to have commitments, long-term projects, etc., in our world. So he would not know what it is like to decide whether he ought to allow people to do certain acts or to let certain states of affairs prevail because these acts and states of affairs affect the commitments and long-term projects that define his very identity. Consequently, Jeeves does not know what it is like to desire for a certain fact not to obtain. His detachment from natural world that he knows about could cause Jeeves to be completely indifferent to what happens in it. Since having a desire for some fact not to obtain is arguably necessary for

11 Pelczar takes this desiderative relation to be the relation crucial for the knowledge that such and such a fact is not morally good. “Necessarily, anyone who knows X is wrong desires that X not be done. ... The claim is not that knowing that X is wrong requires not desiring that X be done. It is that knowing that X is wrong requires desiring that X not be done (p.33).” But Pelczar does not elaborate on what brings about the desire that X not be done, he merely says that when a person who knows all natural facts is cured of moral ignorance she acquires a desiderative relation toward certain natural facts. The new guise hypothesis about Jeeves being proposed here also relies on the necessity of bearing a desiderative relation to natural facts for moral knowledge. But in contrast with the old fact/new relation strategy, what is being proposed here on behalf of the moral naturalist is being grounded on moral agents’ distinct perspective on the natural world.
knowing whether the state of affairs in question is morally good, Jeeves could be morally ignorant.

Maybe Jeeves could be fallible on moral matters although he is infallible when it comes to natural science because the former, unlike the latter, involves a distinct kind of relation towards knowledge of a natural fact. Perhaps Jeeves could only bear such a relation to a natural fact if he lived in the world in the way that we do. If he remains detached from our situation and disengaged with the concerns that come with being in such a situation, then he would not know the desires that drive our judgments, much less the norms that govern these very desires.\textsuperscript{12} It is not just that we desire that genocide, murder, torture for fun, etc., not be perpetrated because it jeopardizes our plans and commitments. We ought to desire that murder, etc., not be perpetrated even if it were not prudential for us to do so. And it must not be possible to know that murder, genocide, torture for fun, etc., are morally wrong without having the requisite desire and understanding the kind of reasoning that governs such desires.\textsuperscript{13}

To clarify matters, compare Jeeves’s judgment of scientific matters based on what he knows compared with his neutrality on moral matters. Natural science involves a normative element because it depends on

\textsuperscript{12} This account of fallible omniscience contradicts the one given in chapter 5. But since the new guise hypothesis about Jeeves is meant to be a competing account of fallible omniscience, if it turns out to be the right one then the world book argument would not be able to establish its antinaturalist conclusion.

\textsuperscript{13} According to Pelczar what a person who knows all natural facts could be ignorant about “concerns facts that it is the business of normative ethics to describe (p. 44).” If it is assumed that some normative ethical theory is true, then it is easy to say what kind of desires Jeeves would have. But if we suspend judgment on competing normative ethical theories and Jeeves is given the requisite kind of experience, it has yet to be shown what desires Jeeves would have and why he would have them. More on this in the next section of this chapter.
experiments, observation and other forms of data gathering performed by fallible and epistemically limited human beings. The norms that scientists follow regulate such activities and promote responsible belief-formation in the scientific community. But these things are superfluous to someone who knows all natural facts. Unlike ordinary scientists, Jeeves can sit on armchair and theorize about the natural world without lifting a finger. He does not have to gather data, design and conduct experiments or even bother about margins of error. Because he knows all natural facts the possibility of being mistaken about some natural matter does not even arise. Jeeves knows all natural facts and all laws of nature. Perhaps his omniscience would even allow him to have the scientist’s practical power of prediction even if he does not move from his armchair. But if Jeeves has never left his armchair the experience of what it is like to be part of the natural world and to desire that this or that state of affairs not obtain would be completely unknown to him.\footnote{Jeeves may have existing desires in his sequestered state, e.g. the desire for a more comfortable armchair, a room with a different view, etc. But having these and other similar desires do not necessarily translate to having desires relevant for knowing whether something is morally good.} He would not know what it is like to be committed to long-term projects, to conceive of himself in terms of valuable experiences, desire that certain states of affairs obtain, etc. And because he does not desire that certain acts be committed or states of affairs come to pass he cannot be in the requisite desiderative relation towards natural facts. If it is right that it is not possible to know whether something is morally good without having such a desire, then moral naturalists won’t have a problem with Jeeves.

Here is the distinct perspective on knowledge of natural facts that the naturalist could use to mount a new guise hypothesis about Jeeves. In ethics, the first person perspective on what ought to be done is a deliverance
of deliberation on what one knows, what one desires and what norms govern these desires. If Jeeves does not know what it is like to take this perspective then it is no wonder that he does not know whether something is morally good. Unlike us, Jeeves has never been in a position in which he has to choose between alternative courses of action. He has never known what it is like to see a set of alternatives from the distinctive first person point of view that common folk find themselves in on a daily basis. Since knowing what it is like to take the first person perspective on what ought to be done about some moral matter is necessary for knowing whether something is morally good, Jeeves would not know whether something is morally good. But this does not mean that such knowledge isn’t already latent in knowledge of all natural facts.

The new guise hypothesis about fallible omniscience being proposed here could be better understood by contrasting Jeeves with Peter Railton’s account of a fully informed moral agent. Unlike Jeeves, Railton’s all knowing moral agent is competent with reasoning that involves beliefs, desires and the norms that govern desires because he begins as an ordinary moral agent who has to make moral judgments using limited knowledge of the natural world. Being a reductive naturalist, Railton believes that the moral goodness of something can be analyzed within existing or prospective empirical theories. He believes that the central evaluative functions of morality are localized in nature and that empirical research can help us uncover which natural properties underwrite these evaluative functions. More

---

15 Railton’s position is partly based on Brandt’s reductive moral naturalism. According to Brandt, something is intrinsically good for a person if and only if it is something he would want for himself if he were fully and vividly fully informed about it.

specifically, Railton's brand of moral naturalism defines goodness in terms of the objective interest of society. To arrive at such a social account Railton begins with the objective interest of an individual, which he defines as follows:

Give to an actual individual A unqualified cognitive and imaginative powers, and full factual and nomological information about his physical and psychological constitution, capacities, circumstances, history, and so on. A will have become A+, who has complete and vivid knowledge of himself and his environment, and whose instrumental rationality is in no way defective. We now ask A+ to tell us not what he currently wants, but what he would want his non-idealized self A to want—or, more generally, to seek—were he to find himself in the actual condition and circumstances of A.¹⁶

According to Railton, what A+ would choose for A is what is in A's objective interest. Railton uses the heuristic tool of conferring full information about himself and his environment on a moral agent to show how knowledge of natural facts bears on knowing what is good for someone. Jeeves differs from Railton's A+ in the following way: A+ knows what it is like to be situated in a world but Jeeves does not. So A+ would know what A ought to do and what he ought to want for himself because he knows not only all the relevant natural facts, he also knows what it is like to be a moral agent. Having had the experience of common folk, A+ knows what it is like to consider different courses of action from the first person point of view. That is why A+ can conceptualize the natural facts that he knows about A in terms of their relevance to what A ought to do.

A moral naturalist like Railton could block the second premise of the world book argument by using these differences between Jeeves and A+ to pose a new guise hypothesis about Jeeves. If Jeeves gains the type of information that A+ has about conceptualizing natural facts from the point of view of a moral agent, then Jeeves could not fail to know whether something is morally good. He wouldn't just know what is good for him or what is good

for A, he would also know what is good for everyone both as individuals and as a collective. The reductive naturalist could then say that fallible omniscience does not imply antinaturalism. All that fallible omniscience implies is that someone can know all natural facts without knowing how to engage in reasoning that involves beliefs, desires, and norms about desires because he has never known what it is like not to be detached and disengaged with the world. But once we give him that experience, the person who knows all natural facts will immediately see what is good for him and what is good for other people. From there, someone whom we might call a non-sequester Jeeves, could easily see what he ought to do when he finds himself in some specific situation. And perhaps if such a person who knows all natural facts had enough time, he could write out what each moral agent ought to do, much like Brink envisioned when he described an omniscient person similar to Jeeves.\footnote{Brink does not agree that moral facts are reducible to natural ones. But according to him someone who knew all natural facts would also know what each person ought to do.}

Michael Smith’s conception of moral facts is similar to the reductive naturalists’ in some respects.\footnote{Michael Smith, “Realism,” \textit{Ethical Theory: An Anthology} (Malden: Blackwell, 2007), pp. 72-6.} Although Smith’s account of moral facts has significant affinities with the reductive naturalist’s position, his moral realism is informed by a conception of philosophical analysis that other moral naturalists
Like Brandt and Railton, Smith grounds his analysis of moral goodness on what a fully informed moral agent would have reason to do if he were ‘well-informed, calm, cool and collected.’ According to Smith, such an analysis has the added advantage of weeding out irrational motives for action, e.g. a desire to drown the crying baby in its bath water to stop it from crying. The demand that we only endorse reasons for action that we would have if we were well-informed and fully rational also allows us to move from what is good for someone to what is good for everyone. Smith believes that through rational argumentation on moral matters that makes an appeal to what we would have reason to do if we were ‘well-informed, calm, cool and collected’ we would eventually arrive at a massive convergence of moral

---

19 Brink characterizes this particular aspect of his disagreement with Smith as a difference of opinion at the metaphilosophical level.

"[W]e disagree about the extent to which it is useful to think of moral, metaethical, and (more generally) philosophical theorizing as conceptual analysis. He is an advocate of this perspective; I am not. I suspect that these metaphilosophical differences influence our rather different view about the most plausible form of ethical naturalism: I favor metaphysical naturalism, whereas Smith rejects metaphysical naturalism in favor of a ‘network analysis’ of moral terms." [“Moral Motivation,” *Ethics* 108 (1997): p. 5.]

20 Ibid., p. 74.

21 Ibid.

Unlike Brandt and Railton, Smith develops his moral realist position as a response to what he calls ‘the moral problem,’ which consists in the following inconsistent triad:

(1) Moral judgments of the form ‘It is right to Φ’ express a subject’s beliefs about an objective matter of fact, a fact about what is right for her to do.
(2) If someone judges that it is right that she Φs then, ceteris paribus, she is motivated to Φ.
(3) An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume’s terms, distinct existences.


Smith solves the moral problem by making a distinction between motives or desires, which are merely occurrent in subjects, and reasons for action, which are considered judgments of moral agents. By so doing Smith allows for rational criticism of desires.
beliefs. Such a convergence would not be forthcoming unless there were moral facts grounded on facts about what we have reason to do. Since facts about what we have reason to do are wholly determined by external circumstances of moral agents these must be objective, even if they are dispositional in nature.

Smith’s account of moral facts can be adopted by moral naturalists like Brandt and Railton to develop an account of natural moral facts that begin from the objective interest of one particular individual. Such a naturalist could say that from this objective interest moral agents can discover through rational argumentation what they all have reason to do if they were fully rational and fully informed about natural facts. If Jeeves were allowed to have the requisite conceptualization of natural facts by being localized in some part of the world at some particular time as a moral agent, then he would begin to see what he and other people have reason to do. From there Jeeves could extrapolate all that he needs to know to be morally omniscient.

Reply to the new guise hypothesis about Jeeves

The proponent of the world book argument could begin a reply to the new guise hypothesis about Jeeves by adopting the strategy used by property dualists or phenomenal realists against the new guise hypothesis about Mary. This strategy consists in saying that conceding that there is some physical information that prerelease Mary lacks that may be parsed in terms of a distinct guise or presentation of natural facts is tantamount to conceding that physicalism about conscious experience is false. According to Stephen L. White, there is an important difference between the Morning Star/

---

22 Ibid., 76.
Evening Star case and the pain/c-fiber firing case. In the former there are clearly two physical properties that correspond to the two descriptions of Venus: “the property of being the last heavenly body visible in the morning and the property of being the first heavenly body visible in the evening.” But in the case of pain, in order to produce an analogue for the two physical properties of Venus that correspond to Hesperus and Phosphorous it has to be supposed that the first order property of being painful (which is identical with a neurophysiological property, say, C-fiber firing) has two second order properties: the mental property in virtue of which it has an a priori connection with the phenomenal character of pain, and the physical property by virtue of which it can be picked out as the neurophysiological property that it is. Hence property dualism emerges at the level of second order properties, despite the physicalist’s attempt to reduce phenomenal properties to physical ones through the old fact/new guise strategy. Pain is different from Venus in that the former is inaccessible outside a first person point of view. So if the physical property of being an instantiation of C-fiber firing does not include the phenomenal character of pain then the latter must be ontologically, and not only epistemically, distinct from the former.

In the same vein, David Chalmers points out that there is an important difference between the a posteriori equivalence between water and H₂O, genes and DNA, etc., on the one hand, and brain states and phenomenal properties on the other. He says that if someone were fully informed about physical facts relevant to water and genes, that person would be able to derive from what she knows that water is H₂O and that genes are DNA.

---


24 Ibid.
To explain genes, we merely have to explain why systems function a certain way in transmitting hereditary characteristics; to explain water, we have to explain why a substance has a certain objective structure and behavior. Given a complete physical description of the world, Mary would be able to deduce all the relevant truths about water and about genes, by deducing which systems have the appropriate structure and function.\textsuperscript{25}

But in the case of phenomenal properties even physicalists like Tye and Horgan concede that prerelease Mary would not know what it is like to see a ripe tomato. So there must be an important difference between knowing phenomenal properties and knowing that two natural kind terms are co-referential. Chalmers adds that if Mary can know all physical facts about color vision and still lack information about the phenomenal character of seeing red from the first person point of view, then there must be some belief about such an experience that Mary lacks. Even if the only belief that prerelease Mary lacks is one in which she knows that she is now having a red experience, the opponent of the new guise hypothesis would turn out to be right about the distinct existence of phenomenal properties. All that phenomenal realism\textsuperscript{26} needs is the admission that there is no a priori entailment from the physical to the phenomenal. Since the new guise hypothesis about Mary concedes that there is ‘merely’ an a posteriori entailment from the physical to the phenomenal when they concede that


\textsuperscript{26} Chalmers characterizes phenomenal realism as follows:

“The view that there are phenomenal properties (or phenomenal qualities, or qualia), properties that type mental states by what it is like to have them, and that phenomenal properties are not conceptually reducible to physical or functional properties (or equivalently, that phenomenal concepts are not reducible to physical or functional concepts). On this view, there are truths about what it is like to be a subject that are not entailed a priori by the physical and functional truth (including the environmental truth) about that subject.” ['The Content and Epistemology of Phenomenal Belief,” \textit{Consciousness: New Philosophical Perspectives} (Oxford: Clarendon, 2003), p. 221.]
prerelease Mary lacks some information about seeing red despite her physical omniscience about color vision, then it seems that they are also committed to admitting that phenomenal properties have a distinct existence from all the physical facts about them. This conclusion is reinforced by the fact that there is a demonstrable difference between the a posteriori connection between phenomenal properties and brain states from an a posteriori equivalence between water and H$_2$O, genes and DNA, Hesperus and Phosphorous, etc.

The property dualist and phenomenal realist strategy against the new guise hypothesis about Mary implies that a similar hypothesis about Jeeves also runs into the same trouble. The moral naturalist proponent of the new guise hypothesis about Jeeves claims that the reason he would fail to know whether something is morally good despite his knowledge of all natural facts is that he has never seen natural facts from the point of view of someone who has to choose what to think or what to do about some moral matter. The sophisticated moral naturalist has it that moral properties are reducible to natural properties associated with what is good for a population or what a fully rational and well-informed collective has reason to do. But the moral naturalist cannot say at the same time that (a) natural moral properties have a distinct mode of presentation unavailable to someone disengaged from the world in the way that sequestered Jeeves is and (b) that once Jeeves is given the requisite kind of first hand experience he will automatically have the right kind of desires, without saying how and why these desires are right. Also, the naturalist has to explain why being non-sequestered could suddenly make Jeeves morally omniscient in the way that favors the naturalist’s understanding of moral omniscience.
To appreciate the naturalist's problem consider this example: before the discovery of DNA, genes were conceived as factors that determine the transmission of traits from one generation to another. Although the exact physical composition of such factors were yet to be known it was hypothesized that the whole gamut of trait inheritance in plants and animals can be attributed to genes that do not blend with one another although they mask one another's effects from time to time. After the discovery of DNA we can say with more confidence that all natural phenomena that could be subsumed under the heading 'heredity' are reducible to biochemical patterns and reactions at the sub-cellular level. Here we have two concepts that turn out to have the same referent and a situation in which such an equivalence allows us to reduce one domain to another. If someone knew all the physical facts about genes, then that person would be able to discover from what he knows that genes are DNA and that hereditary phenomena can be reduced to the action of DNA.

Being fully informed about the natural features of or the physical facts about genes is sufficient for knowledge about DNA and what happens when traits are inherited from parent to offspring. If natural moral properties are discoverable in the same way that we have discovered that the transmission of hereditary characteristics happens through DNA replication, then it is not possible for Jeeves to know what he does and still fail to know whether something is morally good. Being isolated from the experience of common folk would not provide Jeeves with the information that he requires to be morally omniscient. As noted in the previous section, Jeeves's knowledge of all natural facts entails that he would be fully proficient with deriving higher order or more theoretical scientific truths from what he knows even if he does not pass through the practicalities that ordinary scientists do because their
data is limited. So if natural facts are similar to such higher order scientific truths, then it is not possible for Jeeves not to know them. The conceptual difference between natural information as it appears to someone with limited knowledge of natural facts but firsthand experience of making moral judgments, and natural information as it appears to someone who knows all natural facts but does not know what it is like to be a moral agent is not relevant for Jeeves if moral naturalism is true. If moral naturalism is false, then the new guise hypothesis about Jeeves might explain why he does not make any moral judgments.

This initial response to the new guise hypothesis about Jeeves fits well with the difference between ethics and philosophy of mind that was brought up in the previous chapter. Since in the philosophy of mind phenomenal properties are taken to be indubitable, disproving the new guise hypothesis about Mary leads to property dualism or what Chalmers calls phenomenal realism. But in ethics, where the existence of natural moral properties is open to question, fallible omniscience taken together with the demonstration that such a hypothetical being would necessarily be fully informed about a posteriori truths like the co-referentiality of certain natural kind terms and the reducibility of one domain into another implies that moral naturalism is false.

This reply to the new guise hypothesis about Jeeves can be developed further through a reconsideration of one of the reasons given for his moral ignorance. Recall the type of moral disagreement that features well-informed and well-meaning disputants who accept different moral norms. In this type of disagreement the dispute is between two persons or communities who use moral terms to refer to different sets of natural

27 See p. 31.
properties and who subscribe to different normative moral principles. Such a moral disagreement demonstrates that it is possible to know a good number of natural facts about some moral matter, know what it is like to be a person and to have a perspective, desire that this or that state of affairs obtain and still fail agree on the aforementioned moral matter because there are two equally coherent ways of conceptualizing the moral relevance of natural facts. This means that even if it is true that the moral relevance of natural facts could only be properly appreciated when they are seen from a distinct first person perspective, there is no guarantee that all moral agents will see the moral relevance of natural facts in the same way. In Jeeves's case it is not clear what exactly it is that he is supposed to see when he is no longer insulated from our experience of the natural world as moral agents. The moral naturalist cannot maintain that Jeeves will become morally omniscient once he leaves his armchair if she does not show first that the type of moral disagreement described above can be resolved by using only the means available to scientists when they argue about similar matters, or demonstrate how Jeeves would choose between different lives he and all other human beings ought to choose because it is what is good for them. If the moral naturalist cannot do this then it is possible that all that Jeeves will discover when he is released from seclusion is that he has to choose to stick to some established use of moral terms and subscribe to one of the coherent system of moral principles available to moral agents. After all, making such a choice is necessary for survival. But just because Jeeves chooses one way or another does not necessarily mean that his choice has an objective basis.

The problem about what Jeeves would choose to desire is compounded by the fact that it is difficult for someone like him to commensurate all that he knows and find some way to choose between them.
David Sobel and Connie S. Rosati present different but not incompatible arguments for the claim that merely stipulating full-information for a moral agent is insufficient for knowing what is good for such an agent. Without specifying more clearly how such an agent is supposed to commensurate different options in order to determine which one is best for her, it is not clear how she will be able to handle all the information available to her so that one or several possible options trump all the others. Rosati summarizes the problem as follows:

To experience life A as the person she would be in A, a person would need to take on A’s point of view and thus to have certain motivational and cognitive features. To experience life B as she would be in B, a person may need to have different motivational and cognitive features. But the person who is to compare lives A and B will have yet different features. Some of these are cognitive features that correct for or expand her actual cognitive powers. They are thus features she did not have in either life A or B and that would have altered her experience of those lives had she had them. Other features are simply the by-product of the processes she must undergo in order to become fully informed. Yet her features collectively must enable her simultaneously to be informed about what lives A and B were like for her as the persons she was when living them, so that she occupies a point of view that makes each viewpoint accessible. The problem thus more specifically concerns what features might enable a person to be informed in this way.

In other words, merely saying that an agent like Jeeves has cognitive access to all the lives possible for him and for all humans without saying which, if any, of the desires and other motivational features that he has knowledge of ought to guide him is not enough to generate the kind of answer that the moral naturalist like Railton or Smith hopes to find. If A and B were like the participants in the debate on capital punishment described in Chapter 2, it is not clear how Jeeves is supposed to choose which of them is right without

---


assuming that Jeeves would also know that some normative ethical theory is true in the same way that scientific theories are true. But if being omniscient about natural facts is compatible with being morally ignorant then it must be possible for natural facts to underdetermine the truth of normative ethical theories. So the moral naturalist and the moral realist cannot assume at the outset that some normative ethical theory is true in the same way that analogous theories in the natural or social sciences are true.

Sobel raises a separate question about the incoherence of requiring someone like Jeeves to be affected by experiencing the different possibilities that he knows about while at the same time demanding that he remain detached and indifferent towards the different possibilities that he knows about.30 Being affected by the life that one chose to live means identifying oneself with the narrative unity of that life. If Jeeves knows about all such lives but does not identify himself with any of them,31 then he must not know each life in the same way that the persons who live those lives know them. Consequently, Jeeves must not be able choose between these different lives by relying on the motivational features that come with identifying oneself with one of them. But if so, how does Jeeves choose? The question remains unanswered. It seems evident, however, that knowledge of all natural facts underdetermines such a choice. If this is right then the new guise hypothesis about Jeeves poses more problems for the moral naturalist.


31 Both Sobel and Rosati raise questions how, if at all, it is possible to attain this kind of knowledge. For surely knowing what it is like to be sympathetic precludes knowing what it is like to be obtuse (Rosati, p. 318). And knowing that it is like to be Amish is incompatible with knowing what it is like to choose between the many different options that society offers (Sobel, p. 801). So it is not clear what choosing between these options from a fully informed but neutral standpoint amounts to.
Sobel and Rosati’s points about the difficulty of commensurating different options from a neutral standpoint such as Jeeves’s reinforce the problem about finding a way to ground ethical evaluations and imperatives on the descriptive. If one grants that fallible omniscience is plausible it seems that one must also grant that natural facts are neutral between moral value and disvalue. And since there seems to be no non-question-begging way to import normative moral principles onto a neutral, fully informed standpoint the moral naturalist cannot bridge the gap between knowing natural facts and not knowing whether something is morally good as exemplified by Jeeves. Naturalists may yet find some way to bridge this gap, but until they do the world book argument presents a serious problem for them.
I have discussed and defended a new argument against moral naturalism in the present work. Although not entirely unknown to philosophers what I have called the world book argument, which was taken from the only popular lecture delivered by Wittgenstein, is neither widely discussed nor substantially defended. The crucial premise of this argument states that if someone who knows all natural facts does not know whether something is morally good, then moral naturalism must be false. The imaginative possibility contained in this premise is often not taken seriously because of a widespread but mostly unexamined suspicion that it could be explained away easily. One way of looking at my defense of the world book argument against possible objections against it is as an effort to show that a good number of apparently plausible ways of explaining away fallible omniscience do not work.

First, there is the tendency to think that it is only possible to conceive of someone who knows all natural facts but does not know whether something is morally good because of the normative or action-guiding dimension of moral terms and concepts. The discussion of the distinct type of moral disagreement, one in which the participants are well-meaning and well-informed but use moral terms differently, shows that the reason Jeeves is conceivable is not that it is not possible to infer the normative from the descriptive. Rather, the reason Jeeves is conceivable is that there is more than one way to infer moral norms from natural facts that seems to be justified. The relevant type of disagreement shows that it is possible to know the same natural facts about some moral matter but disagree about its moral worth because the participants construe the moral significance of the natural
features of the moral matter in question differently. It is noteworthy that the disputants’ use of moral terms are governed by different semantic norms because of their adherence to different sets of moral norms. One neglected aspect of ethical theory is the relation between the moral norms to which moral agents subscribe and the semantic norms that govern the use of these moral terms. This topic is of course beyond the scope of this work. But it seems to me that some of Wittgenstein’s ideas about what is sometimes called the normativity of meaning and content could be profitably brought to bear on a discussion of the world book argument’s implications for some debates in metaethics. Whatever the outcome of this discussion the attempt to explain away fallible omniscience by citing the normativity of moral terms and concepts is shown to be questionable.

Second, there is the claim that to say that fallible omniscience implies antinaturalism begs the question against the naturalist. Being clearer about what is at stake in the debate between the naturalist and the antinaturalist partly addresses this question. Moral naturalists claim that a certain sort of explanation for moral phenomena can be found. Roughly, they think that massive convergence of opinion on moral matters is possible and forthcoming because the practice of making moral judgments is not dependent on moral instincts that are a mere deliverance of natural selection, historical accident and perhaps the willful intervention of moral agents. Rather, moral phenomena are undergirded by objectively existing natural properties that are analogous to the light reflectance properties of certain surfaces. The outcome of this aspect of the debate between naturalists and antinaturalists depends at least in part on yet to be discovered empirical information on moral disagreement. Until such information becomes available and its philosophical implications for ethical theory are fully worked
out both the naturalist’s and the antinaturalist’s account of moral phenomena stand on more or less equal footing. This is why the world book argument cannot be refuted by saying that fallible omniscience amounts to a petitio.

Third, there is the attempt to draw an analogy between metaethics and philosophy of mind in a way that justifies the existence of natural moral properties. The crucial move against this line of thinking is to show an important disanalogy between moral properties and phenomenal properties. Whereas it cannot be doubted that there is such a thing as knowledge of what it is like to see a colored object, to feel pain, to smell pungent odors, etc., it is not clear whether there really is such a thing as knowledge of natural moral properties. Since the existence of moral properties is open to question, an argument propelled by a character who is fully informed about natural facts yet deficient in moral knowledge would yield an antirealist metaphysical conclusion. Unlike the knowledge argument, which establishes the discrete metaphysical existence of phenomenal properties, the world book argument implies that there might be no such thing as moral properties. Hence the attempt to explain away fallible omniscience by drawing an analogy between phenomenal properties and moral properties fails.

Refuting the world book argument by drawing an analogy between the philosophy of mind and metaethics can be parsed in terms of the claim that knowledge whether something is morally good is already latent in knowledge of all natural facts. But since it is not clear how someone who knows all natural facts can import normative moral principles onto his neutral standpoint in an uncontroversial way, it is difficult to establish that moral knowledge is already latent in knowledge of natural facts. It seems that doing the latter requires commensurating the different points of view of varied moral agents from an impersonal and perspectiveless standpoint; and this sounds
impossible if not altogether contradictory. So again, the attempt to explain away fallible omniscience fails.

Where does all these leave us? It is clear that in the world book argument we have found an interesting and provocative argument against moral naturalism that deserves serious consideration and further discussion. Moral naturalists may yet find a way to explain away the possibility of knowing all natural facts without knowing whether something is morally good. But in the meantime they cannot rely on mere knowledge of all natural facts to vindicate their conception of moral properties. What is more, naturalists, and perhaps all moral realists, cannot rely on full-information or ideal advisor accounts of moral goodness until some suitable answers to the problems posed by the world book argument are found.

In the foregoing discussion of the new argument against moral naturalism, there are some closely related and interesting philosophical problems that were ignored. For instance, it is not at all clear whether fully informed ideal advisors could influence the evaluative judgments of non-ideal moral agents. In some of his stories Wodehouse explores the relationship between an epistemically limited gentleman and his well-informed butler to humorous effect. In some of the passages from these stories the aforementioned philosophical problem is brought out very well:

I reached for the umbrella and hat, and was heading for the open spaces, when I heard Jeeves give that soft cough of his and, turning saw that a shadow was about to fall on what had been a day of joyous reunion. In the eye which he was fixing on me I detected the aunt-like gleam which always means that he disapproves of something, and when he said in a soupy tone of voice ‘Pardon me, sir, but are you proposing to enter the Ritz hotel in that hat?’ I knew that the time had come when Bertram must show that iron resolution of his which had been so widely publicized. In the manner of head-joy Jeeves is not in tune with modern progressive thought, his attitude being best described, perhaps, as hidebound, and right from the start I had been asking myself what his reaction would be to the blue Alpine hat with the pink feather in it which I had purchased in his absence. Now I knew. I could see at a g. that he
wanted no piece of it.

I, on the other hand, was all for this Alpine lid. I was prepared to concede that it would have been more suitable for rural wear, but against this had to be set the fact that it unquestionably lent a *diablerie* to my appearance, and mine is an appearance that needs all the *diablerie* it can get.¹

Whether rational persuasion could resolve the standoff between (Bertram) Wooster and Jeeves is unclear. It is possible that becoming better informed would not necessarily make Wooster agree with Jeeves. Since it is possible to know a great deal of natural facts but not make any evaluative judgments, it is also unclear whether having an evaluative opinion is a function of knowing more facts. Exploring all these possibilities and the reasons that can be given for them together with their implications for the world book argument and ethical theory, however, requires a separate work.

Bibliography


_____.


Conee, Earl.


Copp, David.


Darwall, Stephen, Allan Gibbard and Peter Railton.


Doris, John M. And Stephen P. Stich.


Feigl, Herbert.


Firth, Roderick.


Frege, Gottlob.


Hacker, P.M.S.


_____.


_____.


