Mind-Body Meets Metaethics: A Moral Concept Strategy

(Forthcoming in *Philosophical Studies*)

Helen Yetter-Chappell   Richard Yetter Chappell*

June 26, 2012

Abstract

The aim of this paper is to assess the relationship between anti-physicalist arguments in the philosophy of mind and anti-naturalist arguments in metaethics, and to show how the literature on the mind-body problem can inform metaethics. Among the questions we will consider are: (1) whether a moral parallel of the knowledge argument can be constructed to create trouble for naturalists, (2) the relationship between such a "Moral Knowledge Argument" and the familiar Open Question Argument, and (3) how naturalists can respond to the Moral Twin Earth argument. We will give particular attention to recent arguments in the philosophy of mind that aim to show that anti-physicalist arguments can be defused by acknowledging a distinctive kind of *conceptual* dualism between the phenomenal and the physical. This tactic for evading anti-physicalist arguments has come to be known as the Phenomenal Concept Strategy. We will propose a metaethical version of this strategy, which we shall call the 'Moral Concept Strategy'. We suggest that the Moral Concept Strategy offers the most promising way out of these anti-naturalist arguments, though significant challenges remain.

1

Introduction

Familiar debates in the philosophy of mind between dualists and physicalists are predicated on an assumption of realism about conscious experiences, the remaining question being whether or not this real phenomenon is purely physical in nature. As the aim of our paper is to explore the connections between these arguments in the philosophy of mind, and the naturalism/non-naturalism debate in metaethics, we will make a parallel assumption of moral realism. That is, we should be understood throughout as focusing on the debate between *cognitivist* naturalists and non-naturalists, mirroring the debate between (realist) physicalists and dualists. We further assume that the epistemic and explanatory gaps that anti-physicalist arguments rely upon should be taken seriously, and likewise in the metaethical domain. So we will focus our attention on synthetic (rather than analytic) naturalist views.

The first half of this paper focuses on exploring and relating anti-naturalist and anti-physicalist arguments in metaethics and philosophy of mind. §1 considers the connections between the Open Question Argument against metaethical naturalism and the Knowledge Argument against physicalism about consciousness, and explain why appeals to standard a posteriori identities (like water = H_2O) cannot help the naturalist. §2 explores a moral version of the Knowledge Argument, and introduces an important distinction between *procedural* and *substantive* idealizations of agents. §3 explains the Moral Twin Earth argument, and the general structure that an adequate

response will need to take. Finally, §4—the heart of the paper—explores how metaethical naturalists might address all these anti-naturalist arguments by developing a moral analogue of the physicalist's Phenomenal Concept Strategy.

1   The Open Question Argument and the Knowledge Argument

On one reading, G.E. Moore's famous Open Question Argument aims to show simply that moral terms cannot be defined in natural (nonmoral) terms. No matter how many non-moral concepts you string together, and how you string them together, it doesn't add up to a moral claim. Here is a recreation of Moore's (1903) argument:

1. The statement "this is <u>something we desire to desire</u>, but it is not good" is not self-contradictory.

2. If (1), then 'good' doesn't mean the same thing as '<u>something we desire to desire</u>'.

3. Therefore, 'good' doesn't mean the same thing as '<u>something we desire to desire</u>'.

This can be thought of as an argument schema, where any proposed definition of the good can be substituted for the underlined phrase. Call this the *Conceptual Open Question Argument*.

This initial argument can be extended to argue that moral *properties* cannot be reduced to natural properties: Identities are either analytic or

3

synthetic. The above argument shows that it is not analytic that moral properties are identical to natural properties. But an identity between moral and natural properties also doesn't function as standard synthetic identities do. In standard cases of synthetic identities, one term refers to whatever satisfies a certain functional role, and the other term picks out something that (as it happens) satisfies that functional role. For example, 'water' refers to whatever satisfies the watery role (clear, odorless...liquid around here). H_2O satisfies that functional role. So, though the terms don't mean the same thing, they co-refer. To adhere to this pattern, moral terms would need to refer to whatever satisfies a certain functional role, and some natural property would need to satisfy that role. But moral terms don't seem to be functional role terms.[1] This undermines the claim that there is a synthetic reduction of moral to natural properties. We might call this the *Ontological Open Question Argument*.

Similar arguments can be constructed in philosophy of mind, to argue against an analytic reduction of the mental and against physicalism:

1. The statement "this state is a state in which R-fibers fire[2], but it isn't

---

[1] Synthetic naturalists, who are the target of this paper, accept this. They hold that knowing how everything in the world functions won't enable one to settle the moral facts. By contrast, complete information of how the world functions enables one to work out the referents of 'water' and other functional role terms. One could deny this asymmetry—see, e.g., Jackson and Pettit (1995); Jackson (2000)—but since we are assuming that the epistemic and explanatory gaps should be taken seriously, such views are not the target of our paper. (Note also that while Jackson and Pettit (1995, 28) seek to address the Open Question Argument by pointing out that on their view it's an a posteriori matter what physical properties *fill* the various moral functional roles, the Open Question Argument could be reapplied to their specifications of the functional roles themselves.)

[2] In a play on C-fibers, we're taking the firing of "R-fibers" to stand in for whatever

a state in which there's red phenomenology" isn't self-contradictory.

2. If (2), then 'red phenomenology' doesn't mean the same thing as '<u>state in which R-fibers fire</u>'.

3. Therefore, 'red phenomenology' doesn't mean the same thing as '<u>state in which R-fibers fire</u>'.

This purports to establish the conceptual conclusion that phenomenal terms can't be defined in physical terms. We can similarly beef up the argument to yield an ontological conclusion. Just as in the moral case above, purported phenomenal-physical identities don't function as standard synthetic identities (as we discuss below).

Many well-known arguments against physicalism are closely related to the Open Question Argument(s). Consider the Knowledge Argument (Jackson 1982): Mary is a brilliant color scientist, who's been raised from birth in a black and white room. She has never had a color experience. Despite this, Mary has devoted her life to the study of physics and neurology, and—through her black and white textbooks and computer—she has come to know every general physical fact about light and color vision processing. Now imagine that we present Mary with a beautiful red rose. Intuitively, Mary will be surprised. She will come to think that her previous conception of the world was impoverished. She knew everything that normal people's brains did when they saw red roses, but she didn't know that they had *an experience*

---

actually goes on in our brains when we have red experiences.

5

like this. We can now construct the following argument:

1. Captive Mary knows all the physical facts about color and color processing.

2. Mary learns something when she sees red for the first time (what it's like to see red).

3. So there was some fact Captive Mary didn't know (namely, the fact that such-and-such is what it's like to see red, that such-and-such is the experience normal observers have had all along when they saw red objects).

4. But Captive Mary knew all the physical facts.

5. So the fact Captive Mary was ignorant of was a non-physical fact.

6. So there are non-physical facts.

Like the Ontological Open Question argument, the Knowledge Argument aims to establish ontological dualism—the moral equivalent of which would be non-naturalism. As we will see, some physicalists have argued that it does not succeed in meeting this goal, but rather establishes the weaker thesis of *conceptual dualism.*

Physicalists have given numerous different responses to the knowledge argument. Some—like Daniel Dennett (1991)—simply deny that Mary would learn anything at all. If she really knew all the physical truths to begin with,

they insist, she wouldn't be in the least surprised when she saw red for the first time. It would be just as she imagined it. Proponents of the Ability Hypothesis argue that Mary would learn something, but what she gains is simply knowledge-how: a new ability to distinguish red things from non-red ones in a way she couldn't before. But many find these responses unpalatable. It seems crazy to deny that Mary would be surprised by her first glimpse of redness. And she intuitively learns more than just a new way of identifying red things. She seems to gain the propositional knowledge that *this* is what ordinary observers experience when they look at red objects.

Because of this, the most popular strategy among physicalists is to accept that Mary does learn a fact, but to insist that this was not a *new* fact, but rather an "old fact" under a new guise. The physical facts are *all* the facts. And since Mary knew all the physical facts to start, there was no fact that she was ignorant of. But there are different ways of *describing* these facts. Mary may have known all of the facts even before she had a red experience, but she didn't know them under all descriptions. In particular, she lacked the phenomenal ways of thinking about these facts until she had her first experience of redness.

Proponents of this strategy originally tried to defend it by appeal to standard cases of necessary a posteriori truths. Water is necessarily H_2O, though we can only come to know this fact a posteriori. Similarly, it was argued, the phenomenal-physical truths are necessary, but can only be known a posteriori. When you learn that water is H_2O, you are not learning a new

fact over and above the fact that water is water. You are simply learning this old fact in a new way. Likewise, Mary is simply learning an old physical fact under a new phenomenal guise.[3]

However, most philosophers of mind now agree that this appeal to standard cases of necessary a posteriori truths is not a promising way to defend their view. Mary can know all of the physical truths, but still be ignorant of the phenomenal truths. Imagine a parallel case in which George knows absolutely all of the microphysical truths. If George knows all of the microphysical truths, he will be in a position to work out (without any further investigation of the world) that gloms of H_2O molecules behave in certain ways that result in their being a clear liquid, boiling at 100 C, freezing at 0 C, and so on. George will be able to infer that H_2O actually fulfills the watery role. And because actually fulfilling the watery role is what it takes to be water, he will be able to infer that water is H_2O. If the phenomenal-physical truths followed the model of standard necessary a posteriori truths, Mary— with her complete physical knowledge—should be in a position to infer the

---

[3] This same analogy to standard a posteriori identities has been—and continues to be— used by metaethicists in responding to the Open Question Argument. For example, Schroeder (2007, 72) writes:

> Moore's Open Question argument...is hard to take seriously, once we clarify that reduction is a metaphysical thesis, rather than one about our normative concepts....All of the Open Question tests for cognitive significance distinguish between Hesperus and Phosphorus...But no one concludes that Hesperus is not Phosphorus.

The thought is that goodness *just is* desire satisfaction (say), even though we can only come to know this identity fact a posteriori, just as in the case of water-H_2O and Hesperus-Phosphorus.

phenomenal-physical truths. But she is not (Jackson 2004).

Because of this, physicalists attracted to the old fact, new guise approach have more recently opted to appeal to the special nature of phenomenal concepts to explain Mary's predicament. There are two ways phenomenal concepts might be relevant to this argument. We'll introduce the first—what we'll call the *New Concept for Mary Strategy*—simply in order to distinguish it from the more promising *Phenomenal Concept Strategy*. Once we've laid out the responses to the Knowledge Argument, we will see whether we can construct a moral parallel of the Knowledge Argument and whether the same sorts of responses are promising in replying to it.

First, you might explain why Mary learns something when she sees red for the first time as follows: Captive Mary had all the relevant physical concepts. But she didn't have any phenomenal concepts, as phenomenal concept possession requires having had the relevant phenomenal experience. Mary learns something in (2) because she acquires a new concept: a phenomenal concept. Call this the *New Concept for Mary Strategy*.

Most philosophers of mind agree that the New Concept for Mary Strategy is not sufficient to defuse the Knowledge Argument, as we can create a stronger version of the argument in which Mary has all the relevant concepts, but she still seems to learn something new. Suppose that we go into Captive Mary's room, and we show her a red piece of paper. But we don't tell Mary what color the paper is.[4] It seems Mary can now form a concept of the ex-

---

[4] Note that Mary is only understood to have *general* physical knowledge. Her background knowledge does not include specific facts like the reflectance properties of the particular

perience she's having as she looks at the paper (a phenomenal concept). But though she has the relevant phenomenal concept and she has all the relevant physical information, there *still* seems to be something she is missing. This is evidenced by the fact that Mary seems to learn something when we tell her "A ripe tomato is *this* color." (Stoljar 2005)

For this reason, philosophers of mind have recently proposed another way of making sense of the knowledge argument as an argument for conceptual dualism. The idea is that phenomenal concepts and physical concepts are conceptually isolated, such that Mary could know all the physical facts and have all the phenomenal concepts and yet still be ignorant of the phenomenal-physical truths. Hence, when we show Mary a paradigmatically red object, or point to something red and tell her "that's red", she gains knowledge: the kind of knowledge that comes from merging two distinct conceptual apparatuses.[5] This view is known as the *Phenomenal Concept Strategy*.[6] Defenders of the Phenomenal Concept Strategy fill the picture in with a theory of phenomenal concepts designed to show why phenomenal concepts are conceptually isolated in this way. Numerous different theories have been proposed

---

piece of paper we brought into her room.

[5] As, e.g., when someone with the concept BARACK OBAMA and the perceptual concept THAT GUY realizes that that guy is Barack Obama. See Perry (2003). (Though the knowledge that Mary acquires is general knowledge about what red looks like, rather than merely particular knowledge about the object in question. She isn't just attributing a property to a particular object, rather, she's merging her perceptual and descriptive-scientific ways of knowing about *redness*.)

[6] Even if you think that the Knowledge Argument succeeds in demonstrating the stronger conclusion of ontological dualism, it presumably still requires a conceptual dualism between the phenomenal and the physical.

toward this end: indexical theories (Perry 2003), direct reference theories (Levin 2006; Tye 2003), and constitutional theories (Papineau 2002; Balog 2012a).

2 Knowledge Argument and Moral Concepts

How does the Knowledge Argument—and these strategies for showing its force to be merely conceptual—relate to moral knowledge?

If we understand the Knowledge Argument as suggested by the New Concept for Mary Strategy, we can create a straightforward moral parallel. We might imagine a creature who has complete knowledge of the natural (non-moral) truths, but who lacks moral concepts. Suppose we somehow bestow upon her moral concepts and thereby enable her to see actions *as right*. She might seem to learn something. As in the phenomenal case, this would not yet show that there's a new *property* that she's become acquainted with, but simply that she's acquired a new conceptual framework. The Open Question Argument makes plausible that some creature could be conceptually deficient in this way (as a creature with just enough to understand the non-moral descriptions, but not whatever extra element we have that enables us to have a concept of goodness).

But we saw that the New Concept for Mary Strategy does not dispel the mystery of the phenomenal Knowledge Argument, since Mary seems to learn something even when she has all the relevant concepts. Is the same true of the parallel moral case? Would "Moral-Mary", once in possession of moral

concepts, still be ignorant of the moral truths?

There is an argument to suggest that she would not be morally ignorant. Mary is supposed to be ideally rational, able to work out all the implications of her background knowledge. (It is of no interest that some *irrational* agent might know the natural truths but fail to know the moral truths, any more than that some irrational agent might fail to appreciate that the microphysical facts entail that water is H_2O.) And the fundamental moral principles are widely thought to be a priori, assuming that they are knowable at all. An ideally rational agent would presumably (be in a position to) know all the a priori truths. Hence, since Moral-Mary knows all the natural truths, she would would (be in a position to) know the fundamental moral truths. Give Moral-Mary the relevant moral concepts and she won't be ignorant of anything![7]

But one might feel that there is some sleight of hand involved in this argument. We are interested in whether Moral-Mary *merely* acquires a new concept, or whether she moreover acquires new propositional knowledge (not analytically equivalent to any of her past knowledge). Moral knowledge, unlike phenomenal knowledge, is plausibly a priori (at least in the weak sense that one may justifiably believe the fundamental moral truths without

---

[7] This isn't quite right. Even if Moral-Mary would be competent in applying her moral concepts from the moment she woke after the operation, she will not yet be able to *identify* her moral concepts with non-moral concepts. This is in common with the case of phenomenal-physical identifications, and in contrast with cases like the identification of water and H_2O. If someone knew all the H_2O facts, but lacked a concept of water, all we would need to do was give them a concept of water to enable them to derive that water=H_2O.

requiring empirical evidence).[8] Because of this, it is not enough for us to ask whether an "ideally rational" Moral-Mary (given moral concepts) would immediately know the moral facts. A positive answer there is compatible with the claim that Moral-Mary gained new substantial knowledge, requiring rational insights of a kind that go beyond the mere application of a new conceptual apparatus to an old domain of facts.

The problem here arises from the fact that "ideally rational" is ambiguous between substantive and merely procedural idealizations. In order for Moral-Mary to be able to work out all of the moral-natural truths as we imagined above, we had to assume that she was ideally *substantively* rational, so that she could work out all the a priori truths: synthetic as well as analytic.[9] She could then plausibly work out the moral truths, showing that the moral facts are not, in this sense, "*rationally* isolated" from the natural truths—they can be bridged by a substantively rational agent. But given that we must leave open the possibility that moral truths are synthetic a priori,[10] we need a different test to see whether Moral-Mary has acquired new synthetic knowledge (over and above merely acquiring a new concept, and the concomitant conceptual truths).[11]

---

[8] Note that this is compatible with Copp (2003)'s rejection of empirically indefeasible or "strongly a priori" moral knowledge.

[9] This substantively rational agent will thus be one who has the *substantively correct* rational intuitions.

[10] Again, in the sense of being substantive truths that can be justifiably believed in the absence of empirical evidence.

[11] It's interesting to note that because the moral truths are not "rationally isolated" from the natural truths (as the phenomenal truths are from the physical truths), we cannot construct a moral parallel of the conceivability argument. This is particularly interesting

To do this, we might instead ask whether a merely *procedurally* rational Moral-Mary, fully conceptually and logically competent, but possessing no further rational insight, would thereby be in a position to identify the moral truths. And here the answer seems to be "No"—there are plausibly many different internally-coherent moral viewpoints that conceptually competent and empirically-informed agents might hold. Merely possessing moral concepts, and knowing all the (non-moral) natural truths, will not allow Moral-Mary to decide between (e.g.) Consequentialism and Deontology. So the moral truths are at least *conceptually* isolated from the natural truths, in the sense that mere conceptual competence is insufficient to bridge the two domains.

In the phenomenal case, rational isolation and conceptual isolation go

---

because in the philosophy of mind, the Knowledge Argument and Conceivability Arguments are taken to be closely related. The fact that they do not similarly go together in metaethics highlights the importance of the distinction between *rational* and *conceptual* isolation. (A related contrast: metaethical non-naturalists agree that the moral facts supervene on the natural facts, whereas mind-body dualists deny that the phenomenal metaphysically supervenes on the physical.) According to the conceivability argument:

1. A fully rational agent with complete physical (i.e. non-phenomenal) information can conceive of a world physically just like ours, but with no conscious experiences.

2. Ideal conceivability entails metaphysical possibility.

3. So such a world is metaphysically possible.

4. But then it's metaphysically possible for a world to be physically just like ours, but lack phenomenal experiences.

5. So physicalism is false: the physical facts do not exhaust the facts.

Consider a straightforward moral adaptation. Could a fully rational agent, with complete non-moral information conceive of a world that's just like ours in all non-moral respects, but differs from ours morally (perhaps in which nothing is right/wrong/ good/bad/etc.)? The answer seems to be 'no'. For this to be conceivable, moral concepts would need to be rationally isolated. And they are not. So the "Moral Conceivability Argument" doesn't get off the ground.

hand in hand, making them easy to conflate. Reflection on this moral parallel shows that these varieties of isolation are distinct and can come apart.

We've argued that the New Concept for Mary Strategy, despite initially appearing more promising in the moral case, ultimately also fails once we make the necessary adjustments in our tests for whether Mary has acquired substantial new knowledge. We saw (i) that procedural rather than substantive idealization seemed most appropriate for testing whether Moral-Mary gains any substantive new knowledge, and (ii) that a merely procedurally idealized Moral-Mary, even given moral concepts, would not yet possess moral knowledge. So, as in the phenomenal case, some further explanation of Moral-Mary's new knowledge is required.

3   Moral Twin Earth

The notion that Moral-Mary would know all the moral truths can be generalized as the thesis that suitably idealized agents would converge on the moral truth, and have no irresolvable moral disagreements. Call this the *Convergence Thesis*. Interestingly, this is just the thesis that naturalists need to defend in order to escape Horgan and Timmons' powerful "Moral Twin Earth" objection. In this section we will explain the objection, show how the Convergence Thesis could allow the naturalist to respond, and—finally—show how standard naturalist views are in tension with the Convergence Thesis. The next section will introduce a new strategy for defending metaethical naturalism by resolving this tension with the Convergence Thesis.

15

Horgan and Timmons (1992) suggest the following argument against metaethical naturalism: First, note that the naturalist is committed to there being some semantic story about how the reference of our moral terms gets fixed. For example, perhaps 'right' and 'wrong' refer to those natural properties of actions that causally regulate our practices of praise and blame. So, if the consequentialist property of *maximizing happiness* is what causally regulates our praising practices, then 'right' will refer to the natural property of an act's being happiness-maximizing.

Second, Horgan and Timmons point out that we can imagine a society very similar to ours but where the "active ingredients" in the moral semantic story are slightly different, such that they end up picking out a different natural property. So, in our above example, we might imagine a world much like ours except that the deontological property of *conforming to the categorical imperative* is what causally regulates our counterparts' practices of praise and blame. So, in that world, 'right' will refer to the property of conforming to the categorical imperative.

This naturalist theory implies that we are talking past each other, both speaking the truth in our own moral language, when we affirm consequentialism and our Moral Twin Earth counterparts affirm Kantianism. This seems an unacceptable relativistic result, and violates our semantic intuition that the two parties are—despite their different answers—addressing the same moral question. Intuitively, we are *disagreeing* with our Kantian counterparts, not merely speaking past one another. (Contrast the standard case

of water/H_2O: In regular Twin Earth, we have no semantic intuition that speakers genuinely disagree when we say "water is H_2O" and they say "water is not H_2O". The standard Kripke/Putnam intuition is that the two parties are talking about different substances. This difference strongly suggests that it would be a mistake to model our moral semantics on the semantics of natural kind terms.)

Horgan and Timmons further hypothesize that this result can be generalized to any semantic story the naturalist might offer about how the reference of our moral terms gets fixed. Whatever the details of the reference-fixing story, they argue that it will be possible to construct an alternative "Twin" world where the same reference-fixing story picks out a *different* property than it does in the actual world.

Perhaps the most promising way for naturalists to respond (in the spirit of Merli 2002) is to preclude the possibility of divergent moral reference by way of the following two claims: (1) The right reference-fixing story appeals not just to our actual (possibly irrational) theories or practices, but rather to an idealized version thereof. (2) All possible agents, when suitably idealized, would converge on the same moral theories or practices.

But this raises the question of whether the appropriate idealization is substantive or merely procedural in nature, posing a dilemma for the naturalist. Any purely procedural process of idealization is too weak to secure moral convergence amongst all possible agents. There are surely multiple possible internally-coherent moral views, any one of which might be endorsed by

agents engaging in wide reflective equilibrium, depending on their starting points.

The naturalist might instead turn to a more substantive idealization. But now the naturalist faces the challenge that they have no basis for claiming that any particular one of the competing, internally coherent moral theories is the *one true* moral theory. After all, given the natural (non-moral) parity between us and our Moral Twin Earth counterparts, what in the two worlds can the naturalist appeal to as the basis for a moral or rational asymmetry between us?

So the naturalist seems unable to appeal to the above strategy of securing moral convergence by way of idealization. Procedurally idealized agents are not guaranteed to converge in their moral judgments. And substantive idealization presupposes what the naturalist has yet to explain, namely how there can be a single true referent for moral discourse when there are multiple candidates all equally eligible from a naturalistic standpoint.

In the next section we'll suggest a way for the naturalist to avoid both the Moral Twin Earth dilemma and the Moral Knowledge Argument, by developing a moral parallel of the Phenomenal Concept Strategy.

4   A Moral Concept Strategy?

Recall that the Phenomenal Concept Strategy is a way of defending physicalism in the philosophy of mind, while granting that Mary learns something when she is released from the black and white room and finds that red ex-

periences *look like this.* According to the phenomenal concept strategy, phenomenal concepts and physical concepts are conceptually isolated, such that even a fully rational agent with complete physical knowledge and all of the relevant concepts could still be ignorant of the phenomenal-physical truths. But although there's a *conceptual* dualism that prevents captive Mary from working out the phenomenal-physical truths, proponents of the phenomenal concept strategy insist that this does not entail an ontological dualism. The phenomenal properties just are physical properties, but there are two ways of conceptualizing these properties that cannot—because of the nature of these concepts—be inferentially bridged.

Perhaps naturalists in metaethics could appeal to a similar strategy. Here is the basic idea: The moral properties just are natural properties. But there's an unbridgeable conceptual gap between the moral concepts and the natural concepts. This gap arises because the two kinds of concepts are radically different in nature. Because of this difference, a conceptually competent agent (even with all of the relevant concepts and natural knowledge) can't derive the moral-natural truths. Moral and natural concepts are each indispensable for understanding the relevant domains of discourse. Talk of 'right', 'wrong', 'permissible' and so on cannot be reduced to talk in natural terms. But the fact that moral discourse requires non-natural *concepts* does not entail that it refers to non-natural *properties*.

Importantly, the proponent of this Moral Concept Strategy rejects appeal to standard Kripkean cases of the necessary a posteriori. There are no ideal

19

explanatory gaps in standard cases of necessary a posteriori truths, so we cannot give the standard explanation of the a posteriori status of putative moral-natural identities. Instead we need to tell a distinctive story about moral concepts that will account for our inability to derive the moral-natural truths. This commitment is what sets the Moral Concept Strategy apart.

Let's see how this view allows the naturalist to address the anti-naturalist arguments discussed so far.

The Moral Concept Strategist responds to the Conceptual Open Question Argument and the Moral Knowledge Argument by asserting the existence of a conceptual gap: The conceptually competent agent is not necessarily able to derive the moral truths from the natural truths. This is why claims relating moral and natural facts have a substantive, 'open' feel, and why a conceptually competent Moral-Mary may lack moral knowledge (even when in possession of moral concepts). Importantly, this posited conceptual gap is not just another standard case of the necessary a posteriori—there is no rigidified functional role associated with moral terms, unlike 'water' and other natural kind terms. Because of this, even complete natural knowledge won't put one in a position to identify the referents of the two concepts (unlike water=H_2O). Instead, in the moral-natural case, the Moral Concept Strategist holds that there is a completely independent way of latching onto the moral properties, which doesn't allow us to create such an explanation linking the moral with the natural. This contrast between the moral-natural conceptual gap and standard a posteriori identities also provides a response

to the Ontological Open Question Argument. This argument demanded an alternative explanation of synthetic identities, given that moral terms aren't functional role terms. And this alternative approach is precisely what the Moral Concept Strategy embraces.

The Moral Knowledge Argument had us inhabit Mary's perspective, asking from within whether she learns something substantive and new. Similarly, in the Open Question Argument, we ask ourselves the question "Does such-and-such seem to me to have an 'open' feel?" In these cases we're dealing with "from-the inside" intuitions about moral phenomenology, which seem to reveal a gap between our thoughts about the moral and the natural. The way for the naturalist to accommodate these intuitions is to appeal to a conceptual gap. This is precisely the move the Moral Concept Strategist makes.

The Moral Twin Earth argument requires a different approach. Here the challenge is not to accommodate a perceived conceptual gap, but to bridge an apparent referential (and ontological) gap. The naturalist needs a moral semantics that secures convergence of reference between agents with different moral beliefs, so that they may qualify as genuinely disagreeing over a shared subject matter. We saw that the naturalist's best hope was to appeal to a reference-fixing story concerning agents' *idealized* judgments, given some appropriate idealization that would secure convergence in moral beliefs. But this requires a substantive, morally-loaded idealization, which may seem unavailable to the naturalist, given the natural symmetry between

different internally-coherent moral views.

The Moral Concept Strategist has a response. Following the model of Katalin Balog's (2012b) response to critics of the Phenomenal Concept Strategy, the Moral Concept Strategist can appeal again to the conceptual dualism that is fundamental to their view. Yes, when we use our *natural* concepts we can see no asymmetry between rival moral positions, but it's central to the view that our natural concepts do not suffice to reveal the full picture. To understand moral truths, we must use moral concepts—even though the moral facts just are natural facts. While some may find this puzzling, the Moral Concept Strategist will here thump the table and insist that this is an important commitment of their view. If you take seriously the claim that there are two unbridgeable and explanatorily incomplete ways of conceptualizing the world—using natural concepts and using moral concepts—this is simply what falls out of that picture.

So the Moral Concept Strategy answers the Moral Twin Earth argument by appealing to a moral semantics given in normative terms that create an asymmetry between correct and incorrect moral viewpoints. The posited conceptual dualism allows the naturalist to make this move even though no asymmetry is apparent when the viewpoints are described in purely natural terms. To spell out how this might work, suppose that the reference of moral terms is fixed by the moral judgments of substantively rational agents. The normative property of being substantively rational *just is* some natural property, though it is not transparent which natural property it is. Sup-

pose it is the property of having a utilitarian mindset. Then it will turn out that 'moral rightness' (in both our mouths and those of our confused Kantian twins) refers to the natural property of maximizing happiness, allowing for shared discourse and avoiding the semantic relativism threatened by the Moral Twin Earth argument.

However, one might wonder why we can't run the Moral Twin Earth argument one level up. One might think that the naturalist is committed to there being a purely naturalistic explanation of why 'substantively rational' refers to having a utilitarian mindset, rather than some other (internally coherent) candidate mindset. Couldn't we then construct a 'Twin' world where the explanatory features are tweaked so as to yield (were that world actual) a different account of substantive rationality?

The Moral Concept Strategist will respond just as before. Moral concepts, on their view, are essential for understanding moral truths. Because of this, we cannot give an explanation of why the substantively rational mindset is *this* one rather than *that* one in purely naturalistic terms. But this doesn't mean that the Moral Concept Strategist is rejecting naturalism. The dualism that they embrace is merely a *conceptual* one, albeit a conceptual dualism that runs all the way down. To think about first-order moral truths, we must use moral concepts, and likewise, to think about how our moral concepts refer or why they refer to what they do, we must use moral concepts. At every level, we must use moral concepts if we want to understand moral phenomena. We can restate the metaphysical basis of those phenomena

using merely natural terms, but in doing so we will, according to the Moral Concept Strategy, lose important epistemic and explanatory virtues.

This proposal bears some similarities to Sayre-McCord (1997)'s "moral kinds" response to Moral Twin Earth. On Sayre-McCord's view, moral theory may carve up the world in a very different way from scientific theory, such that our moral talk may pick out a shared moral kind even if causally regulated by disparate natural/scientific kinds (284-5). Like the Moral Concept Strategist, Sayre-McCord rejects the standard water-H_2O analogy in favor of positing a more radical 'conceptual isolation' between the normative and natural domains. Perhaps most significantly, he also sees first-order moral theory as *explanatorily prior* to moral semantics. As Sayre-McCord puts it, "what a moral term refers to, if anything, is determined by whether, in light of the best moral theory, the use of that term can be seen as appropriately regulated by instances of a normatively significant kind." (291)

However, an unfortunate consequence of retaining such a 'kind'-based semantics is that Sayre-McCord is limited in the amount of genuine moral disagreement he is able to accommodate. This is because the kind-based approach explains our possession of a kind concept through our usage being causally regulated by the kind in question (270). Thus Sayre-McCord writes (290):

> If we discovered of a community that their use of the terms 'right' and 'good' were not appropriately regulated by what is right or good but instead by something else we would again have grounds

24

for thinking that they were not using the terms to say of things what we say with ours—even if their terms played a role in guiding their actions.

Sayre-McCord's view could be thought of as one version of the Moral Concept Strategy—a version on which moral concepts are given a kind-based semantics. This view imposes the constraint that our 'goodness'-talk must be regulated by the things that really are good, in order for us possess the concept of *goodness*. But the Moral Concept Strategy could just as well be developed using a different theory of moral concepts. For instance, one view of moral concepts might have them be individuated by a combination of action-guiding role and the phenomenal characters associated with feelings of guilt, obligation, righteous anger, etc. The same moral concept could then be possessed by pairs of people who differ radically in their (even procedurally idealized) dispositions to apply the concept to various worldly items. One of them would be radically mistaken, as a matter of substantive moral fact, but such substantive moral error does not preclude them from have a conceptual grasp of goodness or to-be-pursuedness.

Conclusion

While not ourselves naturalists, we think that the Moral Concept Strategy offers a promising new avenue for naturalists to explore in responding to anti-naturalist arguments. What we've offered in this paper is merely a

sketch of this strategy, but we've seen that it has the potential to provide straightforward responses to the Open Question and Moral Knowledge arguments, and can even answer the Moral Twin Earth argument. One curious feature of the resulting view is that it is a form of naturalism which takes the moral truths to be explanatorily prior to the moral semantics: We refer to such-and-such natural property *because* that's what rightness is, rather than the moral-natural identity holding because our moral concepts refer (in some independently verifiable way) to that natural property. This feature may not appeal to 'Hard Naturalists' who take non-natural concepts to be dispensable, but it should—like the Phenomenal Concept Strategy in philosophy of mind—appeal to those naturalists who respect the intuitive pull of anti-naturalist arguments. The remaining challenge for the naturalist who wants to adopt this strategy is to flesh it out by providing an account of moral concepts that will vindicate the claim of an unbridgeable conceptual gap.

References

Balog, Katalin. 2012a. "Acquaintance and the Mind-Body Problem." In Christopher Hill and Simone Gozzano (eds.), *New Perspectives on Type Identity: The Mental and the Physical*. Cambridge University Press.

—. 2012b. "In Defense of the Phenomenal Concept Strategy." *Philosophy and Phenomenological Research* 84.

Copp, David. 2003. "Why Naturalism?" *Ethical Theory and Moral Practice* 6:179–200.

Dennett, Daniel. 1991. *Consciousness Explained*. Boston: Little, Brown and Co.

Horgan, Terrence and Timmons, Mark. 1992. "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived." *Philosophical Papers* 21:153–175.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32:127–136.

—. 2000. *From Metaphysics To Ethics: A Defence of Conceptual Analysis*. Oxford University Press.

—. 2004. "Postscript." In Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar (eds.), *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, chapter 18. MIT Press.

Jackson, Frank and Pettit, Philip. 1995. "Moral Functionalism and Moral Motivation." *Philosophical Quarterly* 45:20–40.

Levin, Janet. 2006. "What is a Phenomenal Concept?" In Torin Alter and Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism.* Oxford University Press.

Merli, David. 2002. "Return to Moral Twin Earth." *Canadian Journal of Philosophy* 32:207–240.

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

Papineau, David. 2002. *Thinking About Consciousness.* Oxford University Press.

Perry, John. 2003. *Knowledge, Possibility, and Consciousness.* MIT Press.

Sayre-McCord, Geoffrey. 1997. "'Good' on Twin Earth." *Philosophical Issues* 8:267–292.

Schroeder, Mark. 2007. *Slaves of the Passions.* Oxford: Oxford University Press.

Stoljar, Daniel. 2005. "Physicalism and Phenomenal Concepts." *Mind & Language* 20:469–494.

Tye, Michael. 2003. "A Theory of Phenomenal Concepts." In Anthony O'Hear (ed.), *Minds and Persons.* Cambridge University Press.