

BRIE GERTLER

A DEFENSE OF THE KNOWLEDGE ARGUMENT

(Received in revised form 19 June 1997)

Paul Churchland distinguishes two approaches to the study of mind.

[On] “the top-down approach” . . . one starts with our current understanding of what intelligent creatures do, and then asks what sort of underlying operations could possibly produce or account for such cognitive activities. In sharp contrast, [methodological materialism] starts at the opposite end of the spectrum, and is called the “*bottom-up* approach”. The basic idea is that cognitive activities are ultimately just activities of the nervous system; and if one wants to understand the activities of the nervous system, then the best way to gain that understanding is to examine the nervous system itself. . . .¹

Reductive materialists take the “top-down” approach: beginning with our mental concepts, they attempt to discover physical (or functional) states or processes to which the mental can be reduced. Although eliminativists take the “bottom-up” approach, that approach need not yield eliminativism. Whether it does so depends on whether neuroscientific research reveals properties which satisfy our mental concepts.

My project in this paper is to call into question the viability of the prevailing view of the phenomenal, which weds a “top-down” approach to the mind with a materialist ontology. Call this position “top-down physicalism”. I revisit the Knowledge Argument, which aims to show that there is information about the phenomenal which is not reducible to, nor even inferrable from, information about the physical.² Many, like David Lewis, believe that this “Hypothesis of Phenomenal Information” (or “HPI”) threatens physicalism, for it entails that there is a sphere of non-physical facts.³ Others, such as Michael Tye, maintain that physicalism is compatible with HPI, and that the Knowledge Argument relies on illicitly applying Leibniz’s Law within an intensional context. Corresponding to these views about HPI, there are two chief strategies for blocking the Knowledge Argument: (1) analyzing the apparent possession of phenomenal information as the having of an ability, and (2) construing it

as knowledge of facts which are ontologically reducible to (though conceptually distinct from) physical facts. I demonstrate that neither of these strategies succeeds. Top-down physicalism about the phenomenal thus appears untenable; physicalists can then justify their position only by adopting the more extreme, bottom-up methodology.⁴

1. LEWIS'S TWO "UNPALATABLE" CONSEQUENCES

Lewis specifically targets Frank Jackson's Knowledge Argument in his attempt to reconcile physicalism with intuitions about the phenomenal. Jackson asks us to imagine Mary, an (otherwise) exceedingly capable neuroscientist who has spent her life in a black and white room, using a black-and-white television set to monitor experiments she conducts. Through her research, Mary has come to know all of the physical facts concerning visual experience, including its functional properties. Jackson claims that when released from her room Mary "will learn something about the world and our visual experience of it."⁵ This shows, according to Jackson, that there is some information which is not physical information and cannot be inferred from physical information alone. Hence physicalism, "the view that a physical theory of nature can fully describe mental activity", is false.⁶ For Jackson, appeals to non-physical *stuff* are not a serious option; he passes off a reference to ectoplasm as "a bit of fun".⁷ Instead, phenomenal properties, or qualia, are "certain *properties* of certain mental states".⁸

Jackson's conclusion is that qualia properties are not identical to any physical (including functional) properties. This is not to say that phenomenal features of mental states are independent of physical features; Jackson allows, at least for the sake of this argument, that the physical fully determines the phenomenal. In the alternative version of the Knowledge Argument Jackson describes Fred, a man who can discriminate two shades of red where the rest of the population sees only a single color. Jackson claims that if Fred's optical system were implanted in someone else's brain, that person would have the relevant qualia. This case illustrates a change in purely physical circumstances that suffices to cause the phenomenal change. If successful, the Knowledge Argument shows that from even exhaus-

tive information about a state's physical features we cannot infer which non-physical features it has. But no results for standing philosophical issues about the ontological or causal independence of the phenomenal follow from Jackson's argument.

Lewis criticizes the Knowledge Argument by criticizing its conclusion, HPI. His strategy is twofold. First, he argues that HPI has two unpalatable consequences. This does not amount to a conclusive refutation of the argument, on his view; rather, these are reasons to be suspicious of HPI, and to avoid it if possible. Then, Lewis offers what he has claimed we need: a way to avoid HPI. This way consists in construing "knowing what it's like" as an ability.

Consequence #1: Imagining Possibilities

Lewis describes the notion of "information" involved in HPI as that of eliminating possibilities, recognizing that certain states of affairs which are (or were) considered possible, are not actual. According to Lewis, one "peculiar" feature of HPI is that a person in Mary's situation cannot imagine any of the relevant possibilities before the experience which results in most of them being eliminated. He does not believe that this inability to imagine the possibilities refutes HPI, but says "it's peculiar enough to suggest that we may somehow have gone astray."⁹ Does this "peculiarity" pose a problem for the Knowledge Argument? Significantly, Jackson denies that his argument requires it. "The Knowledge Argument does not rest on the dubious claim that logically you cannot imagine what sensing red is like unless you have sensed red."¹⁰ If Lewis is to base an objection to the Knowledge Argument on this "peculiarity", then, he must have independent grounds for committing Jackson to the view that the possibilities cannot be envisaged prior to experience.

However, Lewis fails in the following attempt to show that the possibilities cannot be imagined before experience.

When someone doesn't know what it's like to have an experience, where are the alternative open possibilities? I cannot present to myself in thought a range of alternative possibilities about what it might be like to taste Vegemite. That is because I cannot imagine either what it is like to taste Vegemite, or any alternative way that it might be like but in fact isn't. (I could perfectly well imagine that Vegemite tastes just like peanut butter, or something else familiar to me, but let's suppose I've been told authoritatively that this isn't so.) . . . It seems that the alternative possibilities must be unthinkable beforehand; and afterward too, except for the one that turns out to be actualized.¹¹

Following Lewis, let us suppose Mary has been told that red looks different from every color she has ever seen. Lewis provides no reason to conclude from this that she cannot visualize any color which, given her information, could possibly be red. Assuming that Mary's television is a normal black-and-white model, she has seen shades of grey. Perhaps (as a result of reading Hume, say) Mary attempts to envision a missing shade of grey – a shade midway between two shades which are consecutive in the spectrum of greys she has actually experienced. Such an exercise would not result in Mary's visualizing red, of course; but, if successful, it would allow her to formulate one of the alternative possibilities in advance (*contra* Lewis).

Even supposing that Mary is told that red bears absolutely no resemblance to any color she has seen (just as Lewis was told that the taste of Vegemite resembled no more familiar taste), there is no obvious reason for denying that she could experience phenomenal redness, perhaps by using hallucinogenic drugs. If she could have an experiential concept of redness (as, say, "that vivid color I experienced yesterday") then it seems she could imagine redness, though of course not *qua* redness, for she could not know that her experience was a "seeing red" experience. This point underscores the epistemic nature of Jackson's argument. Mary could have concepts of any number of phenomenal qualities, and concepts of every functional or neural state associated with color, without having any reason to believe, of any particular functional or neural state, that it is correlated with any particular phenomenal quality. So possessing these concepts would not constitute knowing what phenomenal features physically- (or functionally-) identified "seeing red" states have.¹² What Lewis considers a "peculiar" feature of HPI is, then, not a feature of HPI at all.

Consequence #2: Epiphenomenalism

If HPI is correct, then there is what Lewis calls "a phenomenal aspect of the world."¹³ Lewis argues that when combined with the correctness of physics, HPI yields epiphenomenalism. As with the "peculiarity" discussed above, Lewis does not claim that epiphenomenalism refutes HPI, but pronounces it "very queer, and repugnant to good sense."¹⁴

It is on only some views about what physics asserts that epiphenomenalism is yielded by the truth of physics combined with HPI. If physics is limited to claiming that all (explainable) physical events can be fully explained by reference solely to physical causes, epiphenomenalism does not follow from the correctness of physics and HPI. For non-physical causes may overdetermine some physical events. The correctness of physics and HPI together yield epiphenomenalism only on a more robust construal of physics, according to which it claims that only physical causes can have physical effects. Against this more robust construal, it could be argued that the proper subject matter of physics is the physical alone, so denying that anything non-physical could have physical effects is beyond its province. The explanatory power of physics does not, then, suffice to generate epiphenomenalism from HPI.¹⁵

Those moved by the exclusion argument, which claims that the causal sufficiency of the physical leaves no room for the mental to causally influence an event, will feel that embracing pervasive overdetermination so as to avoid epiphenomenalism does little to mitigate the alleged repugnancy. This may be true. But it may be possible to avoid epiphenomenalism even without accepting rampant overdetermination. The exclusion argument for epiphenomenalism rests on a view of the relation between mental and physical properties to which opponents of physicalism need not be committed. Among responses to the exclusion argument which reject this view, perhaps the most promising is that offered by Stephen Yablo, who construes the mental-physical relation as the relation of determinables to determinates, thus avoiding both overdetermination and epiphenomenalism.¹⁶ The general strategy of such proposals is to allow that a mental event and a physical event may have partially overlapping causal powers, in which case they do not compete for the title “cause” in the way the exclusion argument envisions. Insofar as it is premature to think that no such proposal will succeed, the question whether HPI yields epiphenomenalism remains open.

These remarks are obviously not intended to prove that epiphenomenalism is not generated by HPI. By providing *prima facie* evidence for HPI, the Knowledge Argument forces us to do one of the following: deny that HPI yields epiphenomenalism; accept epiphenomenalism; or show that the *prima facie* evidence for HPI

is not conclusive evidence for it. While I favor the first option, Jackson accepts epiphenomenalism, offering reasons for thinking it more plausible than many consider it.¹⁷ Lewis chooses the third of the options above, arguing that the Knowledge Argument does not adequately support HPI. He bases this argument on the availability of an alternative account of what Mary learns upon leaving her cell. I now turn to this account, the Ability Hypothesis.

2. THE ABILITY HYPOTHESIS

The Ability Hypothesis says that knowing what it's like to have a particular experience should be construed as "an ability to place oneself, at will, in a state representative of the experience."¹⁸ Lewis endorses Nemirow's claim that this hypothesis should be accepted because it is consistent with physicalism (and so avoids HPI's unpalatable consequences) while explaining everything HPI explains. It explains why "knowing what it's like" cannot be achieved simply through knowledge of objective facts, for placing oneself in a certain state requires a subjective process. Further, it explains why subjective features of experience are linguistically incommunicable: since it is a familiar fact that (many) abilities cannot be conferred through verbal communication alone, this incommunicability is to be expected. As Lewis puts it, "That's why music students have to practice."¹⁹

Earl Conee presents a compelling argument to show that being able to visualize a particular color is neither necessary nor sufficient for knowing what it's like to see that color. It is not necessary, since a subject with "no visual imagination" could nonetheless know what it's like to see red, at least while attentively considering a current experience of that color.²⁰ It is not sufficient, since a subject who has never experienced cherry red, and so does not know what it's like to experience it, could yet know how to imagine cherry red: namely, by visualizing "a hue midway between burgundy red and fire engine red".²¹

Even if Conee's argument refutes the Ability Hypothesis in its current formulation, according to which imaginative ability is central to knowing what it's like, it does not show that *no* version of the Ability Hypothesis could succeed. An ability theorist could respond

to Conee's argument by taking the possession of a different ability to constitute knowing what it's like to see red. Recognitional abilities seem the most promising candidates, especially since the other ability Lewis mentions, the ability to remember the look of red, is subject to the following counter-example: at the initial moment of Mary's first encounter with a red object, she may realize that *this* is what it's like to see red. But since the moment has not yet passed, she is as yet unable to remember what it's like to see red.

Recognitional abilities' intimate connection with knowledge arguably render them immune to this counter-example, as well as to Conee's cases. To exclude Mary's ability to correctly class subjects' "seeing red" states by their physical properties (such as being usually caused by gazing at ripe tomatoes), the relevant ability would have to be more narrow than the ability to recognize "seeing red" states or experiences.²² Suppose, then, that it is the ability to recognize these states *by their phenomenal quality*.²³ To give the ability theorist the best possible case, let us assume that knowledge of what it's like to see red is coextensive with the ability to recognize an experience of seeing red by its phenomenal features – i.e., that anyone who has one of these also has the other. This means that this version of the Ability Hypothesis is not susceptible to a refutation on the model of Conee's, above.

The Ability Hypothesis asserts that the coextensiveness of knowing what it's like and having the recognitional ability is due to the identity of these. Now if knowing what it's like just is having the recognitional ability, then this knowledge can explain the ability in only a trivial sense, if at all. But does the following express a triviality?

- [A] I can recognize seeing-red experiences (by their phenomenal features) *because* I know what it's like to see red.

This statement seems both non-trivial and true. Its non-triviality aside, the distinction between knowing what it's like and the recognitional ability is illustrated by the apparent falsity of the converse claim:

- [B] I know what it's like to see red *because* I can recognize seeing-red experiences (by their phenomenal features).

Only when this “because” is construed evidentially does [B] seem plausible. When we read “because” in its explanatory sense, as we read the “because” in [A], [B] appears false.

In order to show that knowing what it’s like does not consist in a recognitional ability, these two statements need not have different truth values; it is enough if they *may* have different truth values or, to put it slightly differently, if they have different truth *makers*. A detailed examination of Mary’s first sight of a red object illustrates the distinctness of the appropriate recognitional ability, on the one hand, and knowing what it’s like to see red, on the other. Upon her release, Mary lacks both of these. Her ability to recognize “seeing-red” experiences through their *non-phenomenal* features (e.g., being produced in a normal observer when that observer eyes the top part of a traffic light, under normal conditions) leads her to take the experience of gazing at the top of a traffic light as a “seeing-red” experience. By noting the experience’s visual phenomenal quality, Mary learns what it’s like to see red. This learning – correlating a phenomenal feature with “seeing red” – explains her subsequent ability to recognize “seeing red” experiences by their phenomenal features.

At the very least, it is plausible to think that learning what it’s like explains Mary’s subsequent ability. But no plausible competing explanation is generated by reversing the direction of explanation, to suppose that having the appropriate recognitional ability explains Mary’s knowing what it’s like. (The recognitional ability which partially explains how Mary comes to know what it’s like is the inappropriate one described above, which Mary had prior to her release.) For a recognitional ability is the ability to apply a standard correctly, and that presupposes having learned which standard is the correct one – in this case, having learned which phenomenal feature “seeing red” experiences have, i.e., what it’s like to see red.

The fact that one of these is much more plausible, as an explanation, than is the other, implies that [A] and [B] have different truth makers. This implies, in turn, that knowing what it’s like does not consist in being able to recognize red. So the alternative version of the Ability Hypothesis also fails.

An objection to this line of argument charges that the propositional knowledge analysis of knowing what it’s like is built into the

above description of Mary's release. I concede that the above account treats knowing what it's like as propositional knowledge, but I maintain that this functions as a conclusion rather than as a premise of the argument. Ability theorists can reject this account in one of two ways: they can take Mary's new ability to recognize "seeing-red" experiences as a brute ability; or they can offer an alternative account of her acquiring that ability. The former is clearly inadequate – to say that she gains this ability only after she is released and sees something red implies that there is some story to tell about *how* experiencing red effects this change in her. To be adequate, any such story must say what it is about having the experience that enables Mary to recognize other such experiences through their phenomenal features. The obvious answer is that having the experience allows Mary to correlate "seeing red" with the relevant phenomenal quality – in other words, that having the experience makes her realize that "seeing red" experiences have *this* phenomenal quality. This is to attribute Mary's ability to the propositional knowledge of phenomenal qualities ability theorists sought to avoid; but there seems to be no room here for an alternative account.

Moreover, this explication of Mary's learning process explains the appeal of an ability analysis. Aiming to formulate an ability coextensive with knowing what it's like, so as to avoid a Cone-style refutation, we have arrived at an ability the possession of which closely approximates the having of propositional knowledge about the phenomenal quality. "*S* knows that seeing red is like *this*" strongly implies "if *S* underwent an experience with *this* phenomenal feature, she could recognize it (by its phenomenal quality) as red". And the converse is also true.²⁴ This should come as no surprise, since "recognize" is an epistemic term. In its use here, to recognize seeing-red experiences is to class such experiences under the concept "seeing red".

Opponents of HPI may welcome this result, taking it to justify recasting, as statements concerning abilities, statements which purport to ascribe propositional knowledge of phenomenal features, and thereby to license rejecting the alleged propositional knowledge which they find theoretically suspect. The appeal of an ability analysis likely derives from the close tie between propositional knowledge about a phenomenal quality and the ability to recog-

nize it. After all, it is the *consequences* of HPI which physicalists find distasteful; and one way to avoid these consequences, while acknowledging the intuitive force of HPI, is to reject propositional knowledge of the phenomenal in favor of something closely tied to it. As the above argument demonstrated, however, mere coextensiveness does not suffice for reduction, let alone eliminative reduction. So even if ability theorists could define an ability that was coextensive, in the above sense, with knowing what it's like, or with what HPI's defenders claim is propositional knowledge, this would not justify rejecting HPI. In the case we considered, the details about how the ability results from experience yielded *support* for HPI.

The failure of the Ability Hypothesis means that Lewis's rejection of HPI rests entirely on the value he finds in avoiding the epiphenomenalist consequences he attributes to it. Lewis shares the intuition that Mary learns something about color experience upon her release; but without a tenable alternative account of what she learns, he can remain loyal to this intuition only by either accepting epiphenomenalism or denying that it follows from HPI. Although Lewis has illustrated the importance, for physicalism, of finding a way to sidestep HPI, he has not succeeded in rescuing physicalism by finding that way.

An adequate criticism of HPI requires, then, another account. We now turn to the second top-down physicalist strategy, pursued by Conee and Tye.

3. CONEE'S ACQUAINTANCE ACCOUNT

Conee construes as "knowledge by acquaintance" what pre-release Mary lacks regarding phenomenal features. He argues that this construal affords a way to accept that Mary learns something upon her release while denying that she learns any facts.²⁵

Conee uses the following two premises to show that knowledge of phenomenal qualities requires acquaintance: (1) "Having knowledge of any sort implies achieving some optimal cognitive accomplishment with reference to the object of knowledge"; and (2) "experiencing a quality is the most direct [and so the optimal] way to apprehend the quality."²⁶

Statement (1) appears vulnerable to counter-examples. E.g., I can come to know that it is raining if someone who is in a position to

know, and who has always proven trustworthy, tells me that it is. Yet any “cognitive accomplishment” of mine is certainly not “optimal”: I have neither looked out the window nor ventured outdoors. Similarly, (1) seems to imply that we do not gain knowledge through reading about events we might have experienced first-hand.

Even allowing (1) and (2), this account leaves unanswered the most pressing question: Why is it that, in contrast to other objects of knowledge, we can know (or effect an “optimal cognitive accomplishment” with respect to) phenomenal qualities only by experiencing them? Conee is silent as to whether it is unique to phenomenal qualities that apprehension through representations does not suffice for knowledge of them. Yet he takes this requirement to distinguish the phenomenal from uncontroversially physical qualities such as those possessed by cities, bicycles, and umbrellas. Premises (1) and (2) suggest that this disparity is explained by our ability directly to apprehend phenomenal qualities, while we can apprehend cities, etc., only through representations. Yet the central question still remains: What is it about phenomenal qualities that underwrites this epistemic difference?

Conee’s acquaintance view emphasizes the epistemic disparity between uncontroversially physical features (including being a certain neurophysical state) and phenomenal features, but does nothing to explain this disparity. All explanation must end somewhere; but Conee’s fails to start. This lack of explanatory power does not characterize acquaintance analyses, generally – for example, Russell posited reified sense-data to explain our acquaintance with the phenomenal, thereby treating the phenomenal as ontologically distinct from the physical. It is the combination of the acquaintance view with the denial that the epistemic difference derives from an ontological difference which makes Conee’s account unsatisfactory.

The “directness” Conee mentions is a promising step toward an explanation; completing that explanation would require filling out this notion in a way that makes clear why a difference in such “directness” yields an epistemic difference. The Knowledge Argument suggests that phenomenal qualities are essentially experiential, and in this way ontologically different from the properties of cities, bicycles, and umbrellas. This ontological difference could explain a difference in the relation to the knower’s mental states which might

be termed a difference in “directness”. So the ontological difference implied by HPI is a substantive explanatory hypothesis. By attributing the phenomenal’s epistemic specialness to our ability to apprehend it directly, while using “direct” in a purely epistemic sense, Conee falls into the error made infamous by the case of referring to a substance’s “dormative powers” to explain why it puts one to sleep.

4. TYE’S “NO NEW FACTS” OBJECTION

In his 1986 article, Tye argues that phenomenal knowledge is not knowledge of any non-physical facts, but simply a “new way” to grasp a physical fact.²⁷ His argument embodies a leading objection to the Knowledge Argument: that it illicitly applies Leibniz’s Law within an intensional context. This objection claims that a phenomenal concept and a physical (functional) concept may differ in intension, in how they pick out their referents, while they share the same extension.²⁸ So upon her release Mary learns the intension of “seeing-red”, but she discovers no new features of the world, for she was familiar with the referent of this concept – the physical or functional state – all along.

Tye substitutes for Mary the brilliant twenty-third century scientist Jones, who is blind from birth but will eventually acquire eyesight through an operation. Tye argues as follows for his view that blind Jones is not factually ignorant.

One of the things Jones knows of [a particular experience] *e* is that it has the phenomenal content which is typically caused in [a subject] Smith by his viewing red objects. Jones is also aware of Smith’s rigid name, ‘*R*’, and he knows a description which fixes its referent, namely ‘the phenomenal content which is typically caused in Smith by his viewing red objects’. Hence, Jones knows that *R* is the phenomenal content which . . . etc. Hence, Jones must surely know of *e* that it has *R*. . . . Jones will certainly learn something about *e* when he gains his sight. He will learn or discover what *e* was (phenomenally) like. This is new knowledge, knowledge of one particular experience, but it is *not* knowledge of any new facts.²⁹

Tye’s rejection of Fregean methods of fact individuation is here in evidence. To give Tye the strongest possible case, let us accept, for now, that after acquiring sight Jones learns no new facts.³⁰

The following analogy offers a rough measure of the epistemic progress that can be achieved by gaining knowledge of the sort Tye

denies is knowledge of a new fact. A game show contestant must choose between the prize behind curtain #1 and the prize behind curtain #2. Monty Hall tells her that the prize behind the first curtain is an *A* (using “*A*” as a rigid designator), while behind curtain #2 there is a prize of type *B*. The contestant’s situation vis-à-vis the prizes parallels blind Jones’s relation to the phenomenal content of Smith’s experience: just as Jones knows of Smith’s experience that it has *R*, the contestant knows of the prize behind curtain #1 that it is an *A*; just as Jones knows that “*R*” is a rigid name for the phenomenal content Smith typically has under certain circumstances, the contestant knows that “*A*” is a rigid name for the type of thing behind curtain #1. Obviously, the contestant’s hesitation does not show that she is fickle. (Imagine Hall saying, “You know what’s behind each curtain – just tell me which you’d prefer!”) So knowing a rigid name for a property or type, and knowing a description which fixes the referent of the name, leaves plenty of room for ignorance about the property or type.

Of course, the game show contestant differs from blind Jones in that her indecision could be cured by learning some physical facts about the prizes, whereas Jones already knows the physical facts concerning Smith’s experience. But this difference does nothing to diminish the force of the analogy. The game show case illustrates the epistemic significance which acquiring what Tye considers a new way of conceiving a previously known fact can have, generally.

Given that so much can be gained by learning a new way to know an old fact, there seems little need to insist that what Jones learns is a new fact. The point of the analogy is to demonstrate that, on this conception of fact, denying that Jones learns any new facts upon gaining his sight does not amount to denying that he makes epistemic progress. So intuitions about the epistemic progress consequent on Jones’s operation are consistent with a denial that he discovers any new facts, in Tye’s sense. It is, then, crucial to find an alternative account of the change in Jones’s (and Mary’s) epistemic situation effected by the relevant phenomenal experiences.

In his recent book, Tye identifies phenomenal qualities with functional properties and attempts to explain why phenomenal qualities are epistemically special. His explanation is similar, in spirit, to Conee’s. As compared to other functional properties, he claims,

phenomenal qualities are special in that possessing a concept of a phenomenal quality requires having (in the case of indexical concepts such as “*this* shade of red”) or having had (in the case of predicative concepts such as “shade of red”) experiences with that quality.³¹ As an explanation of the epistemic specialness of the phenomenal, simply pointing to the role of experience in developing a phenomenal concept does not improve upon Conee’s account. The “explanation” Tye offers merely introduces a new explanandum, one that hardly weakens the threat to physicalism: why does gaining a phenomenal concept require experiencing that quality, while acquiring concepts of ordinary physical things does not require experiencing them?

To show that the conceptual difference between the phenomenal and the functional poses no special explanatory problems, Tye argues that his functional theory of phenomenality parallels the claim that water is H₂O, an identity which, according to Kripke’s persuasive argument, is a metaphysical truth but not a conceptual truth. Tye acknowledges Kripke’s point that the appearance-reality gap which grounds the epistemic possibility that water is not H₂O is not available to explain the apparent possibility of the presence of a particular phenomenal feature without its associated functional property, for when it comes to phenomenal qualities the appearing *just is* the being. Yet, Tye suggests, the fact that this avenue of explanation is closed does not affect the deeper parallel between phenomenal qualities and water, namely that identity statements concerning each may be informative because both are designated rigidly.

Let us distinguish two versions of the thesis that we use rigid designators to refer to phenomenal qualities. The weak version is that our way of referring to phenomenal qualities picks out the same qualities in all possible worlds. This version will not establish that a phenomenal quality is identical to a physical or functional property, for it does not exclude the possibility that the operation of our rigid designator is governed by purely phenomenal essences. In that case, rigidly designated phenomenal qualities could differ, physically and functionally, across worlds.

To show that phenomenal qualities are identical to functional properties, then, Tye must exploit the stronger version of the rigid designation thesis. This version conjoins the weak version with the

claim that cross-world identity is a matter of physical (or functional) identity. The following argument appears the most promising defense of the stronger thesis.

- 1) Phenomenal qualities are natural kinds.
- 2) Empirical science is authoritative as to the essence of natural kinds.
- 3) Empirical science distinguishes kinds by physical (including functional) properties.

Therefore,

- 4) Phenomenal qualities have physical (functional) essences.

Regardless of whether it is premise (1) or premise (2) which bears the primary burden of the argument (and this depends on how we define “natural kinds”), the argument’s force derives from its claim that empirical science, the science of the physical, carries authority regarding the essence of phenomenal qualities. This central claim marks a crucial disanalogy with the water case. The success of that case is due to empirical science’s clear mandate to decide the essence of water. This mandate stems from the conceptual truth that water is identical to *physical stuff of some sort or other*. But no such conceptual truth to the effect that a phenomenal property is identical to some physical (functional) property or other is available to underwrite Tye’s proposed identity. While admitting that a conceptual gap divides the phenomenal and the physical, Tye fails to see that the source of this gap precludes the analogy with water he advances.

Faced with the absence of a conceptual link between the phenomenal and the physical even at the level of generality at which water is linked with “physical stuff of some sort or other”, physicalists may choose to view the authority of empirical science as a valid *prescription*. Its failure as a *description* of our concept of the phenomenal means that we *need* not yield to the deliverances of empirical science regarding the phenomenal, but physicalists can claim that we *should* allow empirical science to arbitrate the phenomenal. Tye may embrace this strategy, for he justifies his functional theory by its explanatory power, and the explanatory power of empirical science is what motivates the physicalist prescription.

Now it is precisely this explanatory power which is at issue. The conceptual gap between the phenomenal and the physical constitutes a deficiency in explanatory power, for one primary goal of explanation is to bridge conceptual gaps. Admittedly, bridging conceptual gaps is not our sole explanatory purpose, and other desiderata are sometimes thought to outweigh this one. The salient alternative desideratum in this case would be *making the phenomenal safe for physicalism*, an end which many see as advancing more common explanatory goals such as simplicity and coherence. Endorsing this as the proper target amounts to prescribing that we deem empirical physicalist science the arbiter of the phenomenal; and the strength of that prescription suffices, as we have seen, to establish the viability of phenomenal-physical (functional) identities.

All of the philosophers discussed here share the intuition that phenomenal qualities are epistemically special compared with physical (functional) properties. This intuition exhibits the conceptual gap separating the phenomenal from “some physical property or other”, which in turn distinguishes phenomenal-functional identities from “water = H₂O”. The physicalist stance regarding the phenomenal is then fundamentally different from that regarding water, for the former is prescriptive while the latter is descriptive. The primary source of support for the physicalist prescription, namely its explanatory power, carries weight only given the basic physicalist commitments. Endorsing this prescription then amounts to what Churchland, in the quote with which this paper began, calls the “bottom-up approach” to the mind: our intuitions about mental properties (the “top”) are valid only insofar as they can be captured by physical properties (the “bottom”).

5. CONCLUSION

The canvassed physicalist accounts of the epistemic progress phenomenal experience can bestow appear wanting. The Ability Hypothesis and the acquaintance analysis Conee formulated to supplant it were shown to be inadequate. Tye’s claim that Mary gains only a new way to know an old fact, and his proposal concerning the phenomenal’s epistemic specialness, also fail to explain her epistemic progress.

The disagreement between physicalists, on the one hand, and those who believe that there is irreducibly phenomenal information which undermines physicalism, on the other, exhibits a tension between two explanatory goals: simplicity and comprehensiveness. While top-down physicalism aims to achieve both of these goals, physicalist accounts of the phenomenal are ultimately driven to sacrifice comprehensiveness for simplicity in a way that effectively dismisses folk concepts and intuitions as unimportant to determining the ontological status of phenomenal properties. Purportedly top-down physicalists then occupy a more extreme position, according to which the capacity to explain intuitions such as those behind the Knowledge Argument is not a standard by which accounts of the mental must be judged. In other words, they effectively become bottom-up physicalists.

Rejecting the physicalist prescription also leaves us with notorious difficulties, of course. Modesty favors remaining agnostic as to whether mind-body relations are forever outside our grasp; at least, taking ourselves to have established such “cognitive closure” puts too much stock in our present cognitive situation.³² The stance of modesty that undermines that pessimistic view also counsels against seemingly optimistic functionalist theories of the phenomenal, for their optimism exacts a steep price, namely, relinquishing any hope of *explaining* key intuitions about the phenomenal. My view about this issue parallels Parfit’s view about ethics. “Some people believe that there cannot be progress in Ethics, since everything has been already said. . . . I believe the opposite.”³³ Whatever the eventual outcome of this debate, it is as yet too early to settle for “explanations” which presuppose that the sort of explanation we really want is unattainable.³⁴

NOTES

¹ Churchland (1984), p. 96.

² Contemporary statements of this challenge to physicalism include Jackson (1982) and Nagel (1974).

³ Lewis (1988). Many physicalists, including Lewis, accept that the hypothesis of phenomenal information is incompatible with physicalism. Support for the view that HPI undermines physicalism can also be found in Kim (1989). Kim argues that nonreductive materialism is not a stable position, for it cannot provide an adequate account of mental causation. Kim’s claim has the consequence that HPI

and physicalism are incompatible, for if the only tenable form of physicalism is reductionism then physicalism entails that there is no irreducibly phenomenal information.

⁴ Bottom-up physicalism is immune to the Knowledge Argument, for the bottom-up approach denies the relevance of the intuitions which fuel it.

⁵ Jackson (1982), p. 471.

⁶ The definition is Laurence Nemirow's. Nemirow (1990), pp. 490–491.

⁷ Jackson (1986), p. 394.

⁸ Jackson (1982), p. 473.

⁹ Lewis (1988), p. 512.

¹⁰ Jackson (1986), p. 392.

¹¹ Lewis (1988), p. 512.

¹² An ordinary person's knowledge of what it's like for someone else to see red does not, of course, derive from an awareness of relations between functional or neural states and phenomenal qualities. Rather, such knowledge proceeds from an implicit belief that when others view an object, they have experiences with phenomenal features similar to those one's own experiences have when one views the object oneself. Because Mary is prevented from viewing colored objects, she cannot use this ordinary method to grasp what it's like to see red; any grasp which she has on what it's like to see red depends on her recognizing correlations between the functional or neural facts about experiences of seeing red and the phenomenal features of such experiences.

¹³ Lewis (1988), p. 507.

¹⁴ *Ibid.*, p. 514.

¹⁵ A consequence of this claim about the subject matter of physics is that physics alone cannot show the physical to be causally closed. The causal closure of the physical is, rather, a metaphysical conclusion.

¹⁶ Yablo (1992).

¹⁷ Jackson (1986), pp. 474–476.

¹⁸ Nemirow (1980), quoted in Lewis (1988), p. 514.

¹⁹ Lewis *ibid.*, p. 516. But note that, while Lewis is correct to point out that abilities often require more than propositional knowledge, it is also true that acquiring propositional knowledge is one standard way to gain new abilities: e.g., when I learn that you live at 14 Main Street, I gain the ability to find your house. Someone who takes the Knowledge Argument to show the truth of HPI can allow that Mary gains all sorts of new abilities upon her release; the 'friend of phenomenal information' (to use a term of Lewis's) simply maintains that Mary's acquisition of some of these abilities is due in part to her gaining propositional knowledge about the phenomenal.

²⁰ Conee (1994), p. 139.

²¹ Conee *ibid.*, p. 138. Conee claims that only the first of these is a counter-example to Lewis's view. On his reading, Lewis construes knowing what it's like to see red as a conjunction of abilities: to remember, imagine, and recognize the look of red. Yet on the page where he expresses that view, Lewis also asserts: 'These abilities to remember and imagine and recognize [the taste of Vegemite] are abilities you cannot gain (unless by super-neurosurgery, or by magic) except by tasting Vegemite and learning what it's like.' Lewis (1988), p. 516. This implies that the second of Conee's counter-examples also constitutes an objection to Lewis.

²² Following Richard Warner, Janet Levin argues that the possibility of having the wider recognitional ability without the narrower one shows that experiencing a phenomenal quality is not necessary for understanding it. Space considerations preclude an adequate treatment of Levin's argument; she thinks that the supporter of HPI improperly constrains what counts as the concept of redness. Warner (1986); Levin (1986).

²³ Still, this may not be precise enough – for while she is in her cell Mary can recognize Joe's 'seeing-red' states by their phenomenal qualities, in some sense, since it is those qualities which lead Joe to say 'I'm seeing red now.' Another difficulty is that abilities are sometimes taken to have different possible timespans than knowledge: though it is ordinarily accepted that someone may realize what it's like to see red at one moment, and suffer total amnesia at the next, momentary abilities may be considered more problematic. I suggest that a subject possesses, at *t*, an ability to recognize red by its phenomenal quality *iff* she is, at *t*, such that if she had an experience of phenomenal redness at that moment, she would (*ceteris paribus*) count it as an experience of phenomenal redness, regardless of whether she had non-introspective sources of information about her experience, e.g., an image of her brain seen through a cerebroscope.

²⁴ If the recognitional ability were appropriately modified, a Nozickian tracking account of knowledge could identify propositional knowledge about the phenomenal with this ability. Such an identification would be fatal to the Ability Hypothesis, since denying that Mary gained propositional knowledge would then be inconsistent with the claim that she acquired this recognitional ability.

²⁵ "There are closely related facts, such as the fact concerning phenomenal redness that red things look *that* way. But we have no reason to doubt that Mary knew all such facts before knowing how red things look. Mary already had the capacity to form thoughts using this demonstrative sort of reference to phenomenal qualities. She was able to demonstrate them with comprehension, at least via others' experiences of them, e.g. as 'that look' while indicating another person's attentive experience of phenomenal redness." Conee (1994), p. 142. The assumption that Mary comprehends the property to which she refers as described begs the question against Jackson's argument. If successful, that argument shows that no non-introspective demonstration essentially manifests a grasp of what it's like to see red.

²⁶ Conee, *ibid.*, p. 144.

²⁷ Tye (1986), p. 9.

²⁸ Proponents of this style of objection to dualist arguments include J.J.C. Smart (1962) and Antoine Arnauld (1984).

²⁹ Tye (1986), pp. 12–13.

³⁰ In Tye (1995), he puts this point by saying that Mary learns "fine-grained", but no "coarse-grained", facts.

³¹ Tye, *ibid.*, Chapter 6.

³² The term is Colin McGinn's. See McGinn (1989).

³³ Parfit (1984), p. 453.

³⁴ For helpful comments on previous versions of this paper, I thank Jaegwon Kim, Charles Raff, Ernie Sosa, Jim Van Cleve, and an anonymous reviewer for this journal.

REFERENCES

- Arnauld, A. (1984): 'Fourth Set of Objections', in Cottingham et. al.
 Chappell, V. (eds.) (1962): *The Philosophy of Mind*, Englewood Cliffs: Prentice Hall.
- Churchland, P.M. (1984): *Matter and Consciousness*, Cambridge: MIT (Bradford).
- Conee, E. (1994): 'Phenomenal Knowledge', *Australasian Journal of Philosophy* 72(2), 136–150.
- Cottingham, S. and Murdoch, trans. (1984): *The Philosophical Writings of Descartes: Volume II*, Cambridge: Cambridge University Press.
- Jackson, F. (1982): 'Epiphenomenal Qualia', *The Philosophical Quarterly* 32, 127–136, reprinted in Lycan (1990), 469–477.
- Jackson, F. (1986): 'What Mary Didn't Know', *The Journal of Philosophy* 83, 291–295, reprinted in Rosenthal (1991), 392–394.
- Kim, J. (1989): 'The Myth of Nonreductive Materialism', *Proceedings and Addresses of the American Philosophical Association* 63, 31–47.
- Levin, J. (1986): 'Could Love Be Like a Heatwave? Physicalism and the Subjective Character of Experience', *Philosophical Studies* 49, 245–261, reprinted in Lycan (1990), 478–490.
- Lewis, D. (1988): 'What Experience Teaches', *Proceedings of the Russellian Society*, University of Sydney; reprinted in Lycan (1990), 499–519.
- Lycan, W.G. (ed.) (1990): *Mind and Cognition*, Oxford: Blackwell.
- McGinn, C. (1989): 'Can We Solve the Mind-Body Problem?', reprinted in McGinn (1991), 1–22.
- McGinn, C. (1991): *The Problem of Consciousness*, MIT Press (Bradford).
- Nagel, T. (1974): 'What is it Like to be a Bat?', *The Philosophical Review* 83, 435–450.
- Nemirow, L. (1980): 'Review of Nagel's *Mortal Questions*', *Philosophical Review* 89, 475–476.
- Nemirow, L. (1990): 'Physicalism and the Cognitive Role of Acquaintance' in Lycan (ed.), 490–499.
- Parfit, D. (1984): *Reasons and Persons*, Oxford: Clarendon Press.
- Rosenthal, D.M. (ed.) (1991): *The Nature of Mind*, Oxford: Oxford University Press.
- Smart, J.J.C. (1959): 'Sensations and Brain Processes', reprinted in Chappell (1962).
- Tye, M. (1986): 'The Subjective Qualities of Experience', *Mind* 95, 1–17.
- Tye, M. (1995): *Ten Problems of Consciousness*, MIT Press (Bradford).
- Warner, R. (1986): 'A Challenge to Physicalism', *Australasian Journal of Philosophy* 64, 249–265.
- Yablo, S. (1992): 'Mental Causation', *The Philosophical Review* 101, 245–280.

Department of Philosophy
College of William and Mary
Williamsburg, VA 23187
 USA