



Inquiry

An Interdisciplinary Journal of Philosophy

ISSN: 0020-174X (Print) 1502-3923 (Online) Journal homepage: <http://www.tandfonline.com/loi/sinq20>

Destabilizing the knowledge argument and modal argument

Amber Ross

To cite this article: Amber Ross (2017): Destabilizing the knowledge argument and modal argument, Inquiry, DOI: [10.1080/0020174X.2017.1385530](https://doi.org/10.1080/0020174X.2017.1385530)

To link to this article: <http://dx.doi.org/10.1080/0020174X.2017.1385530>



Published online: 13 Oct 2017.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Destabilizing the knowledge argument and modal argument

Amber Ross

Department of Philosophy, University of Toronto, Toronto, Canada

ABSTRACT

Several of the most compelling anti-materialist arguments are motivated by the supposed existence of an unbridgeable epistemic gap between first-person subjective knowledge about one's own conscious experience and third-personally acquired knowledge. The two with which this paper is concerned are (i) Frank Jackson's 'knowledge argument' and (ii) David Chalmers's 'modal argument'. The knowledge argument and the modal argument are often taken to function as 'two sides of the same coin ... in principle each succeeds on its own, but in practice they work best in tandem'. This paper disagrees with the above claim, and argues that when considered together the knowledge argument and modal arguments illuminate each other's weaknesses. These weaknesses become apparent when we acknowledge the epistemic richness of the cognitive aspect of perceptual experience, and question the epistemic role that any non-cognitive aspect might play. Closer examination of judgments about what it's like to have a perceptual experience reveals fundamental problems with the knowledge argument, and when these problems are exposed, the two arguments can no longer 'buttress each other where help is needed'. It becomes clear that neither is likely to succeed on its own, and when taken together both are destabilized.

ARTICLE HISTORY Received 12 April 2017; Accepted 21 September 2017

KEYWORDS Perceptual experience; knowledge argument; modal argument; conceivability/possibility principle; consciousness

1. The epistemic richness of the cognitive aspect of perceptual experience

There is a largely underappreciated puzzle concerning the notion of phenomenal belief – a belief about what it's like to have a conscious perceptual experience – and the role it plays in making phenomenal judgments. Making a judgment is a type of action, and according to those who endorse the causal closure of the physical world (materialists and anti-materialists alike)

CONTACT Amber Ross  amber.ross@utoronto.ca

© 2017 Informa UK Limited, trading as Taylor & Francis Group

all actions have fully physical explanations. Thus, there will be a sufficient fully physical explanation of why a subject makes a judgment such as, 'this is a perceptual experience of redness'; an explanation that nowhere requires the subject to have irreducibly phenomenal knowledge. But if judgments about our perceptual experience do not require a subject to have irreducibly phenomenal knowledge, then in principle a subject could identify colors – and perceptual experiences of color – without specifically phenomenal knowledge of those perceptual experiences. Chalmers has labeled this puzzle 'the paradox of phenomenal judgment' (Chalmers 1996, 177ff), and it illuminates the epistemic richness of the cognitive aspect of perceptual experience. Any judgment a subject may make about her phenomenal states requires nothing more than that subject engaging the cognitive aspect of her perceptual experience.

Though Chalmers takes this 'paradox' to be ultimately benign, the following argument will show that, in acknowledging the epistemic richness of the cognitive aspect of perceptual experience, the paradox of phenomenal judgment ultimately destabilizes both the knowledge argument and the modal argument against materialism. If we keep the paradox in mind and consider Mary's situation in the black and white room, we see that her vast physical knowledge guarantees that she will leave the room equipped with the ability to immediately identify her perceptual experiences of color. If this conclusion is correct, it problematizes not only the knowledge argument but the modal argument as well.¹

The tenability of the modal argument essentially depends on the inconceivability of materialism; as such, acknowledging the genuine conceivability of materialism² will destabilize the modal argument. As we will discuss

¹Mary's ability to identify color on sight has consequences for anti-materialism as well as certain materialist theories of consciousness. In addition to problematizing the knowledge argument and modal argument, it weakens the intuitive foundation for 'type-b materialism' or 'thick materialism' (Graham and Horgan 2000). Thick or type-b materialism is the view that while physical properties and phenomenal properties are not metaphysically distinct, there is a clean conceptual divide between physical and phenomenal facts (see Chalmers 2010, 115–118), and that phenomenal facts are essentially involved in explanations of phenomenal knowledge. By treating physical and phenomenal facts as *conceptually isolated*, 'thick' or 'type-b' materialists can explain Mary's inability to infer phenomenal facts from her knowledge of physical facts, while maintaining that all *properties* involved are nonetheless physical properties. If Mary can identify colors on sight solely in virtue of knowing all the physically reducible facts about color and perceptual experiences of color, her ability lends strong credibility to the opposing type-a materialist claim: Mary will know that *having a conscious experience of redness is like this* solely in virtue of knowing all the physical facts there are to know about color and perceptual experiences of color.

²In this paper, I will be presenting a case for the positive conceivability of type-a materialism, rather than making the (potentially easier) case for type-b materialism. This is because type-b materialism is open to certain objections to which type-a materialism is immune; in particular, type-b materialism posits a conceptual gap between the physical and phenomenal, to which anti-materialists may object that this so-called conceptual gap is a metaphysical gap in disguise. According to type-a materialism, there is no gap between the physical and phenomenal, neither conceptual nor metaphysical, and therefore no similar objection can be levied against type-a materialism.

below, the inference from the conceivability of zombies to their possibility only holds so long as materialism fails to be conceivable in the appropriate way. If we can construct a scenario that shows materialism to be conceivable in the appropriate sense, the modal argument will be fundamentally destabilized. 'Anti-zombie' parity arguments expose this weakness in the modal argument but stop short of illustrating a scenario in which materialism is shown to be 'ideally positively conceivable', a scenario the supporter of the modal argument must explicitly deny (see Chalmers 2010, 180). However, if we accept the causal closure of the physical domain and acknowledge the epistemic richness of the cognitive aspect of perceptual experience, we will see that Mary will leave her black and white room capable of identifying colors on sight, which in turn supports the (plausibly ideally) positive conceivability of materialism. Denying that Mary could identify colors on sight upon leaving the black and white room would require either, (1) rejecting the causal closure of the physical, or (2) denying the epistemic richness of the cognitive aspect of perceptual experience. As I will explain, neither option is available to a naturalized property dualist. Without a rigorous argument to show that materialism fails to be ideally positively conceivable, we arrive at this strange but inevitable consequence of the epistemic richness of the cognitive aspect of perceptual experience: it exposes the weakness of the knowledge argument, and in so doing destabilizes the modal argument as well.

2. The modal argument: from zombies to property dualism

Naturalized property dualism assumes that there are psychophysical laws in our world, laws that govern the relationship between the cognitive and phenomenal aspects of our conscious experience such that, as a matter of empirical fact, the phenomenal aspect of our conscious experience *nomologically* supervenes upon our cognitive (i.e., physical) properties. But the dualists argue that the phenomenal facts of our world – facts about the character of our conscious experience – neither *logically* nor *metaphysically* supervene upon physical facts. If phenomenal facts were to supervene upon physical facts in such a way, it would be logically or metaphysically necessary that the physical facts of our world entail the phenomenal facts. Biological facts, for example, logically supervene on physical facts. We could say, following Chalmers (who borrows an image from Kripke 1972), that once God fixed all the physical facts of the world, the biological facts came along for free. But after setting the physical facts in place, God had 'more work to do to make sure there is a law relating the [physical] facts and the

[phenomenal] facts' (Chalmers 1996, 38). According to property dualism, biological, chemical, and *cognitive* facts logically supervene upon physical facts; phenomenal facts stand apart as metaphysically and logically independent of any other kind of fact.

To understand how proponents of the modal argument make their case for the conceivability of zombies, assume for the sake of argument that facts about the character of our conscious experience are not fully determined by facts about our physical nature. If this is possible, then it is conceivable for a subject's physical properties to be exactly as they are now, while her conscious experience differs in some way. It would be conceivable for you to feel a slight ache in your left foot that you are not actually feeling now though your physical body is in a state that is molecule-for-molecule identical to its current state. That *ache* would be a phenomenal property of your experience, and if it is conceivable for there to be a small change such as this in your phenomenal properties, while your physical properties remain the same, philosophical zombies must be conceivable as well: creatures physically identical to human beings but wholly lacking conscious experience. Though the leap is large, the difference is a matter of degree rather than kind. If physical facts do not logically entail phenomenal facts, then it is conceivable for all the physical properties of the world to be just as they are now though the phenomenal properties differ in some respect. And if it is conceivable for phenomenal properties to change or differ without a change or difference in physical properties, it is conceivable that phenomenal properties could be *absent* from a world altogether: a zombie-world.

According to property dualists, so long as zombies are genuinely conceivable they are possible. And if zombies are possible, physicalism is false. If zombies are possible, then it is possible that the phenomenal consciousness we find present in our world – the feature that separates our world from zombie-worlds – could have been *absent*, while the physical facts remained just as they are now. But, according to the property dualist, this is not how our world turned out; our world does contain phenomenal consciousness. If our world is not a zombie-world, our world must contain an additional kind of property from those that exist in zombie-worlds; it must contain both phenomenal and physical properties.

The modal argument may be the most compelling argument for property dualism, and the premises of the modal argument are generally given in this simple form:

- (1) Zombies are conceivable
- (2) Whatever is conceivable is possible
- (3) Zombies are possible
- (4) Materialism is false

But it can be laid out more precisely as follows:

- (1) $P \& \sim Q$ is conceivable.
- (2) If $P \& \sim Q$ is conceivable, $P \& \sim Q$ is metaphysically possible.
- (3) If $P \& \sim Q$ is metaphysically possible, materialism is false.
- (4) Materialism is false. (Chalmers 2010, 142)

As Chalmers articulates the modal argument here, P is ‘the conjunction of all microphysical truths about the universe’, and Q is ‘an arbitrary phenomenal truth’ (Chalmers 2010, 142). A ‘phenomenal truth’ is a particular fact about conscious experience; it can be a fact about an individual’s conscious experience or about phenomenal consciousness in general. So premise 1 of the modal argument states that you can conceive of a world in which the phenomenal facts are not necessitated by the physical facts. Most philosophers accept premise 1; they agree that we can conceive of the physical facts of our world being just as they now and the phenomenal facts differing in some way.

Premise 2 of the modal argument – the controversial premise – is an instance of the Conceivability/Possibility principle (CP). This is the claim that whatever is conceivable is possible, or, more precisely,

CP: Ideal positive conceivability entails metaphysical possibility.³

The distinction between *prima facie* and *ideal* conceivability is intended to separate those scenarios that merely seem conceivable to a subject from those which are actually conceivable. The notion of *prima facie conceivability* is fairly straightforward: S is *prima facie* conceivable for a subject when S is conceivable for that subject on first appearances (Chalmers 2002).

Positive and *negative* conceivability are concerned with *creation* and *elimination*; a scenario is positively conceivable when it can be created in (modal) imagination (Chalmers 2002). A scenario is *prima facie negatively* conceivable

³More precisely still, Chalmers formulates CP within a two-dimensional semantic framework: that ideal primary positive conceivability entails primary possibility. On a 2D semantic framework, a hypothesis will have both a primary and secondary intension, and the conceivability of that hypothesis may depend on whether it is evaluated according to its primary or secondary intension. Though crucial to understanding how CP may stand up to the challenge of ‘Kripkean’ cases of a posteriori necessity, two-dimensional semantics are irrelevant to the challenge to the principle set forth here, and nothing will turn on us leaving aside the notions of primary and secondary conceivability.

when initial consideration of the scenario reveals no obvious conceptual contraction within that scenario. A scenario, *S*, is *ideally negatively* conceivable only if 'the hypothesis expressed by *S* cannot be ruled out a priori even on ideal rational reflection', (Chalmers 2010, 143) and will fail to be ideally negatively conceivable when *S* is found to be *prima facie* conceivable but its *prima facie* conceivability is 'undermined by further reflection showing that the tests that are criterial for conceivability are not in fact passed' (Chalmers 2010, 144).

The notion of forming a positive conception of a scenario, however, is somewhat more complicated, since it involves appealing to one's *imagination*, but not 'imagination' in an ordinary sense; *S* is positively conceivable when one can coherently modally imagine a situation that verifies *S*, which is 'to in some sense imagine a specific configuration of objects and properties', (Chalmers 2010, 145) and 'fill in arbitrary details in the imagined situation such that no contradiction reveals itself' (Chalmers 2010, 145).

Zombies will be *prima facie positively* conceivable when a subject believes that all the details of the zombie-scenario can be filled in without revealing a contradiction, or when a subject believes she can 'imagine a situation with certain important features specified, notes that a situation of this kind appears to verify *S*, and judges that the remaining details are not crucial' (Chalmers 2002, 153).

Of course, mere *prima facie* positive conceivability does not suffice for possibility. 'For the thought-experiment to yield the intended conclusion, this *prima facie* judgment must be correct, so that *S* is ideally positively conceivable' (Chalmers 2002, 154). A subject may make the *prima facie* judgment that she can fill in the details of scenario *S* though the world *S* describes is actually conceptually incoherent.⁴ 'Ideal conceivability' is intended to be treated as a purely rational notion, akin to the notions of *a priority* and *entailment*. A scenario is ideally conceivable if there is no contradiction within it, and it will be ideally *positively* conceivable if a subject could, in principle, 'coherently modally imagine' a world that fits the scenario's description and fill in arbitrary details of that scenario without uncovering any latent conceptual incoherence. Property dualists claim that zombies are ideally positively

⁴Most objections to the conceivability of the zombie-scenario argue against its ideal conceivability, attempting to show that zombies are only *prima facie* positively conceivable, that subjects who believe they are conceiving of a zombie are actually making a mistake of one sort or another in their imagining – possibly by 'filling in' the scenario with improper details, or failing to notice 'holes' in their imagined creature. For arguments of this sort, see Kirk (2005), Dennett on zombies (1991, 2005) and on the knowledge argument (1991), as well as Marcus (2004).

conceivable, and from this infer the possibility of zombies and the failure of materialism .

2.1. CP and anti-zombies

CP is controversial,⁵ and it would be unavailable to the property dualist (without further argument) if a case could be made in which it were shown to imply a contradiction. Several ‘anti-zombie’ parity arguments have recently appeared in the literature, each attempting to use CP and the logic of the zombie modal argument against itself.⁶ Just like philosophical zombies, ‘anti-zombies’ are molecule for molecule duplicates of human beings; the difference between zombies and anti-zombies is that anti-zombies are made conscious by physical facts alone. The proponent of the zombie modal argument claims that zombies are ‘ideally positively conceivable’; that is, if an ideally reasonable epistemic agent were to modally imagine the zombie-scenario, she could fill in all the details of that scenario without encountering a contradiction. Proponents of the anti-zombie modal arguments would like to make the same claim for the anti-zombie scenario: we can modally imagine a scenario in which creatures physically identical to ourselves are made conscious in virtue of their physical properties alone, without contradiction.

For either the zombie or anti-zombie modal arguments to succeed, the possibility of zombies and anti-zombies must follow directly from their conceivability. This is CP at work, and the soundness of the modal argument depends on there being no conceivable scenario in which CP is violated, that is, no scenario that is ideally positively conceivable but metaphysically impossible. This means that if zombies are conceivable, anti-zombies must be inconceivable; one of the two (zombie or anti-zombie) must be merely *prima facie* conceivable if the other is to be ideally conceivable. If one allows

⁵Stephen Yablo gives good reason to reject the controversial CP principle based on the likelihood of human error. Though this is not the argument presented against CP in this paper, it is worth considering in its own right.

If ignorance of an individual’s essential properties can generate modal error, why not ignorance of a property’s essential properties? Imagine that my grasp of a property *S* fails to reflect the fact that it is essentially uninstantiable (*S* might be the property of being sodium-free salt). Nothing to prevent me, then, from conceiving it as possible that *S*s should exist: a *de dicto* conceivability error rather than a *de re* one. Likewise the *de dicto* impossibility that some *Q*s are *R*s will be conceivable, if my understanding of *Q* omits its essential property of having no *R*s in its extension. Probably there is no proposition for which a worry like this cannot be raised. (Yablo 1993, 17)

⁶For examples of anti-zombie modal arguments see Brown (2010), Frankish (2007), Sturgeon (2000), and Marton (1998).

for the conceivability of anti-zombies, the modal argument collapses in contradiction. Assuming that CP holds, then the conceivability of anti-zombies entails their possibility; that is, it would be possible for a creature physically identical to a human to be made conscious by physical facts alone. If materialism and property dualism are both conceivable, and if CP holds, then both zombies and anti-zombies are possible. The result is a contradiction: if anti-zombies are conscious in virtue of the physical facts alone, and the same physical facts hold in the zombie world, then the physical facts of the zombie-world would make the *zombie* conscious. And the concept 'conscious zombie' is straightforwardly self-contradictory (the creature would be, by hypothesis, both conscious and not conscious). Those who would defend the modal argument are faced with a dilemma: they may avoid the problem the anti-zombie poses by either (a) rejecting CP or (b) denying the conceivability of materialism. Given that embracing first horn results in a straightforward rejection of the modal argument, proponents of the zombie modal argument must choose the second.

In defending his anti-zombie argument, Frankish writes, 'To conceive of anti-zombies, we simply have to imagine a world where the relation between a being's phenomenal properties and its underlying microphysical ones is such that the former are not further properties over and above the latter – for example, where the relation is one of token identity' (2007, 656). However, defenders of the zombie modal argument will hold that a materialist theory such as token identity is not obviously coherently conceivable, and in particular the knowledge argument and arguments like it are supposed to make imagining such a case difficult: a subject could know all the microphysical facts but still not know all the phenomenal facts, therefore phenomenal facts must not be contained within the set of physical facts, and token identity fails.

Chalmers explicitly endorses horn (b) of the dilemma above, denying the conceivability of materialism in any philosophically significant way. According to Chalmers, materialism is *prima facie* negatively conceivable at best; that is to say, though materialism may seem to be a coherent theory of mind at first pass, upon rational reflection we will see that materialism is conceptually incoherent. He writes,

It may be *prima facie* negatively conceivable that materialism is true about consciousness, but it [the truth of materialism] is not obviously conceivable in any stronger sense. Many people have noted that it is very hard to imagine that consciousness is a physical process. I do not think that this unimaginability is so obvious that it should be used as a premise in an argument against materialism, but likewise, the imaginability claim [i.e. the imaginability of materialism] cannot

be used as a premise [in an argument for materialism], either. (2010, 180, material in brackets mine)

The ultimate foundation for the claim that zombies are ideally conceivable is the assumption that it is not ideally conceivable for phenomenal facts to logically or metaphysically supervene on physical or physically reducible facts. Chalmers holds that Jackson's knowledge argument in particular should 'compellingly establish the failure of logical supervenience' (Chalmers 1996, 146). If the knowledge argument were to succeed in this task, it would establish the inconceivability of anti-zombies as well. However, in what follows I will show that the epistemic richness of the cognitive aspect of perceptual experience guarantees that Mary will leave her black and white room identifying her perceptual experiences of color on sight. If this is the case, I will argue that the knowledge argument loses its grounds for the claim that Mary learns something new when she sees color for the first time and therefore fails to establish the failure of logical supervenience. Quite to the contrary, Mary's color identification ability illustrates a scenario in which materialism is (potentially ideally) positively conceivable. If materialism is positively conceivable, anti-zombies will be positively conceivable as well, and the modal argument will be in serious jeopardy.

3. Reconsidering the knowledge argument

If Mary can identify her perceptual experiences of colors solely in virtue of the knowledge she gains while in her black and white room, then it is conceivable that Mary had already learned all there is to know about color and the perceptual experience of color before having color experience herself (and without acquiring any *essentially non-physical* phenomenal knowledge). Such a scenario is one in which materialism is (potentially ideally) positively conceivable, and if coherent, it would be sufficient to destabilize the modal argument.

Frank Jackson describes Mary's situation as follows:

Mary is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in completed physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles. (1986, 291)

Mary's scientific color knowledge is exhaustive; she knows all the scientific facts about color and the perceptual experience of color, which is

meant to include all the color facts one can learn without actually having color experience. Subsequent to Jackson's original formulation of the thought experiment, Mary's story has seen extensive revision: her skin has been painted shades of gray, she has been prohibited both from dreaming in color and from rubbing her eyes so as to prevent colorful phosphene experiences, eventually re-created as colorblind from birth. These stipulations have helped to distill the relevant features of her initial state: Mary is supposedly neurologically identical to a normal human being except that she has *not yet* had the experience of seeing color, and she knows all the scientific facts about color and the perceptual experience of color. On some accounts, Mary is simply *scientifically omniscient*; omniscient of all the facts that can be expressed in purely physical and physically reducible terms.⁷

Jackson's argument continues,

If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is what physicalism denies ...

It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described as learning – she will not say, 'ho, hum'. Hence physicalism is false. (1986, 291)

The day that Mary leaves her black and white room, upon having a perceptual experience of her first red object she supposedly has the thought, 'oh, so *that's* what it's like to see red!' As Jackson says, she will not say 'ho, hum'. Her reaction to seeing red is evidence that she learns something upon having her first perceptual experience of color: she learns *what it's like* to have a perceptual experience of redness. There is no dispute over whether Mary entertains *some sort* of novel thought on her first encounter with red; at the very least it has irreducibly indexical content, and Mary has never been in a state about which she could have had such a thought. The relevance of this new thought to the materialism/property dualism debate is that the dualists believe Mary could not have had such a thought without acquiring specifically *phenomenal* knowledge about the perceptual experience of color.⁸

⁷Since all third-person facts (scientific facts) must be physical or physically reducible, Mary may be made omniscient regarding all scientific facts – not merely color facts – without violating any assumptions of the knowledge argument. As Jackson writes, 'Physicalism is not the noncontroversial thesis that the actual world is largely physical, but the challenging thesis that it is entirely physical. This is why physicalists must hold that complete physical knowledge is complete knowledge simpliciter' (1986, 291).

⁸Type-b or thick materialists hold a similar view, although on their account this new knowledge would be knowledge of physical properties conceived under newly acquired phenomenal concepts.

In William Lycan's formalization of the anti-materialist argument within this thought experiment, Mary is color-blind from birth and becomes omniscient regarding all physically reducible scientific facts. Eventually, her color-blindness is cured. Lycan's formulation is as follows:

- (1) Before her cure, Mary knows all the scientific and other 'objective' facts there are to know about color and color vision and color experience, and every other relevant fact. [Stipulation.]
- (2) Upon being cured, Mary learns something, viz., she learns what it's like to experience visual redness. [Seems obvious.]
- ∴ (3) There is a fact, the fact of what it's like to experience visual redness, that Mary knows after her cure but did not know prior to it. [1, 2]
- (4) For any facts: if $F1 = F2$, then anyone who knows $F1$ knows $F2$. [Suppressed; assumes simple factive grammar of 'know'.]
- ∴ (5) There is a fact, that of what it's like, that is distinct from every relevant scientific/'objective' fact. [1,3,4]
- (6) If materialism is true, then every fact about color experience is identical with some physiological, functional, or otherwise scientific/'objective' fact.
- (7) Materialism is not true. [5,6]. (Lycan 2003, 385)

3.1. *The essential implicit assumption: suppressed premise (1.5)*

Premise (2) is the essential premise of the knowledge argument. In this formalization, the justification for premise (2) is its apparent obviousness. But the reason we find it obvious can be captured in an intermediate suppressed premise (1.5), explicitly noting the empirical evidence supporting the claim that Mary learns something new when she steps into the multi-color world.

Suppressed Premise (1.5): When Mary sees color for the first time, she will not be able to identify colors on sight solely in virtue of what she learns before her first perceptual experience of color.

Suppressed premise (1.5) is an implicit assumption, the implicit assumption essential to making premise (2) seem obviously true. When we imagine Mary's response upon seeing red for the first time, we do not imagine her saying 'ho, hum'. We imagine Mary as 'surprised and delighted' to learn that red looks *like this*, or at least uncertain whether *this* or *that* color is red. If suppressed premise (1.5) were false, if Mary could identify colors solely in virtue of what she learns before her cure, premise (2) would no longer seem obviously true.

The essential premise of the knowledge argument is that Mary learns something upon seeing red for the first time, and the evidence supporting this premise is that she will not know the identity of the colors she sees when she leaves her black and white room. If Mary enters the colored world already equipped with the ability to identify colors on sight, one would need to look

elsewhere for evidence that phenomenal facts do not logically supervene on physical facts. If premise (2) cannot be supported by its own obviousness, if it cannot be supported by Mary's reaction to her first perceptual experience of color, the knowledge argument loses its grounds for the claim that a subject could not know everything there is to know about color experience simply in virtue of knowing all the physically reducible facts about color and the perceptual experience of color. And it is not obvious what other sort of evidence would support premise (2) that does not at the same time beg the question against materialism.

The widely received view that Mary will not identify color on sight is based on two assumptions. The first is that for Mary to 'know colors on sight' she must have a special kind of recognitional concept – a phenomenal concept, such as red_{ph} – which she will acquire when she has a conscious perceptual experience of redness and is represented by the term *this* in '*this is what it's like to see red*'. The second assumption is *conceptual isolation*: phenomenal concepts are supposedly isolated from all non-phenomenal concepts. As such, Mary could conceive of red under certain physical and functional concepts in virtue of knowing all the objective, scientific facts about red; however, no depth or breadth of knowledge couched in physical or functional terms could provide Mary with the concept red_{ph} . Mary will not acquire phenomenal concepts by learning all the physical and physically reducible facts there are to know; no level of facility with these kinds of concepts will amount to Mary knowing *what it's like* to see red. Whatever merits the notion of conceptual isolation may have, making it the focus of our analysis of the knowledge argument obscures something far more important to the materialist defense: the cognitive aspect of Mary's perceptual experience is in itself rich enough to allow Mary to identify colors on sight when she leaves her black and white room.

I will argue that we can coherently conceive of Mary knowing everything there is to know about the perceptual experience of color in virtue of both knowing all the relevant physically reducible facts about the perceptual experience of color and having the ability to use this information in the right way. If Mary emerges from her room identifying her perceptual experiences of color with the ease and speed of a seasoned color-perceiver, the Mary scenario become one which demonstrates the positive conceivability of materialism, and plausibly its ideal positive conceivability. If this is the case, then the knowledge argument (and its ilk) can no longer be said to establish

the failure of logical possibility, which in turn causes trouble for the zombie modal argument as well.⁹

4. Positively conceiving of a materialist version of consciousness

4.1. How the epistemic richness of the cognitive aspect of perceptual experience guarantees that Mary will identify color at first sight

When Mary identifies a color she is engaging in a certain behavior, performing a physical act, and acts such as identification and re-identification are subsets of what have been called the ‘easy problems’ of consciousness. The ‘easy problems’ of consciousness differ from the ‘hard problem’ insofar as explanations of easy problems will only require discoveries of new physical or functional facts within science-as-we-know-it (see Chalmers 1999). As we mentioned at the outset, those who accept the causal closure of the physical world must acknowledge the explanatory irrelevance of phenomenal consciousness to resolving any of the ‘easy problems’ of consciousness: there will be a sufficient explanation of a subject’s every action or behavior that nowhere requires the introduction of new phenomenal properties or phenomenal knowledge but only ordinary cognitive, functional, physical facts. As Chalmers explains,

⁹Daniel Dennett also rejects suppressed premise (1.5) in his (1991) response to the knowledge argument, where he proposes an alternative ending to the Mary story.

And so, one day, Mary’s captors decided it was time for her to see colors. As a trick, they prepared a bright blue banana to present as her first color experience ever. Mary took one look at it and said, ‘Hey! You tried to trick me! Bananas are yellow, but this one is Blue!’ (399)

Dennett claims Mary will know that the banana is blue on sight, but underestimates the strength of his position. He writes,

My variant was intended to bring out the fact that, absent any persuasive argument that this could not be how Mary would respond, my telling of the tale had the same status as Jackson’s: two little fantasies pulling in opposite directions, neither with any demonstrated authority. (2007, 16)

This alternative ending to Mary’s story is presented as one ‘fantasy’ in competition with the received view, neither with any authority. I disagree with Dennett, and maintain that Dennett’s variant of the thought experiment is the only interpretation compatible with the claim that the universe is causally closed under the laws of physics, and its ‘demonstrated authority’ is provided by the paradox of phenomenal judgment. To put the point another way, the knowledge argument is critically weakened not because we *can* reject suppressed premise 1.5; so long as we accept the causal closure of the physical domain we *must* reject suppressed premise 1.5. The assumption that secures the knowledge argument’s essential premise is simply incompatible with the widely held principle of causal closure, a principle few are willing to reject in favor of the prima facie compelling knowledge argument.

Chalmers entertains the possibility that Mary could have such an ability, but does not believe it ultimately undermines the knowledge argument.

Dennett (1991) ... notes that Mary could use her neurophysiological knowledge to recognize that a red object is red when she sees it ... Perhaps this is so, but all that follows is that *contra* Lewis and Nemirow, Mary had certain abilities to recognize even before she had her first experience of red. (Chalmers 1996, 145)

This response seriously underestimates the consequences of Mary’s ‘recognitional abilities’. In addition to categorically undermining the knowledge argument, it illustrates a scenario which, if true, demonstrates the (potentially ideal) positive conceivability of materialism. If so, then what appears to be merely a problem for the knowledge argument, leaving its conclusion contentious at best, threatens to undermine the modal argument as well.

... phenomenal judgments lie in the domain of psychology and in principle should be reductively explainable by the usual methods of cognitive science. There should be a [fully] physical or functional explanation of why we are disposed to make the claims about consciousness that we do, for instance, and of how we make the *judgments* we do about conscious experience. It then follows that our claims and judgments about consciousness can be explained in terms quite independent of consciousness. More strongly, it seems that consciousness is *explanatorily irrelevant* to our claims and judgments about consciousness. This result I call the *paradox of phenomenal judgment*. (1996, 177, emphasis original)

During her time in the black and white room, Mary learns all the physically reducible facts about color and the perceptual experience of color, which includes all the facts there are to know about the cognitive, functional, and other physically reducible states that observers enter into when they perceive a particular color. Her knowledge will include knowledge of all the facts that supervene upon physical facts; the only knowledge unavailable to her is whatever knowledge might be irreducibly non-physical – irreducibly phenomenal facts about the perceptual experience of color, if such facts exist. According to the anti-materialist who accepts the causal closure of the physical world, facts about cognitive states and behavior can be given entirely physically reducible explanations, since phenomenal facts and phenomenal knowledge are explanatorily irrelevant to our understanding of behavior.¹⁰ If Mary is a normal color perceiver, then in learning all the physically reducible facts there are to know about the cognitive states connected to normal subjects' perceptual experience of color, she is learning all the physically reducible facts there are to know about the cognitive states she would be in if she were to have a perceptual experience of color.¹¹

¹⁰These physically-reducible explanations would be as accurate and complete for a human subject as they would for her phenomenally deprived zombie-twin, including explanations of her reactions to color stimuli (see McGeer [2003, 386] and Chalmers [1996, 181]).

¹¹Mary's visual normalcy is not essential to this argument. We typically assume that Mary is a normal color perceiver, so that when Mary does have her first color experience she learns not only what it's like for her to have a perceptual experience of color but facts about what it's like for normal perceivers to experience color. In learning what it's like for normal perceivers to have a color experience, Mary learns a general fact about the world that she could not learn by knowing all the physical facts alone (see Jackson [1986]). This avoids the problem of the indexical content of Mary's new knowledge: before Mary experienced color, there were no facts about Mary's-color-experience for her to know, since no Mary-color-experience had existed while she was in the black and white room. There were facts about what normal color-perceivers experienced, and in learning this fact she learns something that was true before she experienced color but that she did not know.

However, the present argument that Mary will identify color on sight does not depend on the stipulation that Mary is a normal color-perceiver; Mary will learn facts about normal color perceivers, and if Mary is a color-invert (for example, experiencing color *x* as redness while normal perceivers experience *x* as greenness) she will make inverted color identification. Since Mary is actually identifying her color *experience*, not color as a property of an object, Mary will be making an accurate identification though her experience does not map onto objective color properties in the normal way.

4.1.1. 'Highly Attuned' Mary – a positively conceivable materialist scenario

If the knowledge argument is to have any bearing on the viability of materialism, we cannot conceive of Mary as *simply* omniscient regarding the physical facts of her world. We must also stipulate that Mary has the extraordinary but conceivable ability to process this vast information without error. This is first essential feature of the ability we will call being 'Highly Attuned'. Mary must be the kind of epistemic agent who can immediately recognize every relevant feature of each physical fact, within each situation. If Mary were to lack such an ability, she might fail to realize that, for example, *that* experience is a perceptual experience of red', simply because she has failed to notice the conceptual connection between *that perceptual experience of color* and her physically reducible color knowledge. This kind of failure would not indicate a gap between physically reducible color facts and phenomenal color facts; rather, it would merely reveal a weakness in her ability to see the conceptual connection between physical and phenomenal facts. If this sort of fallibility were the source of her error, her error would have no relevance to the tenability of materialism.

Unless we stipulate that Mary is Highly Attuned, we cannot assume that any shortfall in her knowledge of color experience implies that there is some fact or information about the perceptual experience of color that Mary did not have in virtue of knowing all the physical facts alone. Those facts might have been derivable from her vast physical knowledge, yet her cognitive limitations barred her from properly processing the relevant facts at the appropriate time. We cannot, therefore, conceive of Mary's cognitive abilities on the model of an ordinary-though-exceptionally-brilliant epistemic agent who also knows all the physical facts there are to know; such an agent would not be cognitively equipped to make use of the vast body of information that would constitute 'knowing all the physical facts'.

If 'knowing all the physical facts' is to pull any weight in the knowledge argument it must be this sort of useful knowing: Mary's cognitive faculties must be as powerful as her knowledge base is expansive. Any normalcy constraint on Mary's cognitive capacity would weaken the argumentative strength of the knowledge argument. If Mary cannot process all the information provided by her scientific omniscience, her situation would be functionally equivalent to one in which she does not know all the physical facts. And if Mary does not know all the physical facts, then her inability to identify color on sight would be no point in favor of the anti-materialist. Therefore, we must stipulate that Mary learns the distinction between every cognitive state related to the perceptual experience of color to a level as

fine-grained as the minimum distinguishable difference¹² between every humanly distinguishable color.

Of course, ordinary human subjects are in fact highly fallible when judging their own mental states, but this is independent of the *conceivability* of a subject correctly identifying her own cognitive states by directly attending to them. And if it is logically possible for a subject to possess such an ability, then we ought to stipulate that Mary has this ability. This is the second essential element of being Highly Attuned: the extraordinarily refined ability to attend to the minimum distinguishable differences between one's own cognitive states. Being Highly Attuned, then, would involve a highly acute, *strictly cognitive*, non-phenomenal form of introspection.¹³ If Mary is Highly Attuned, she can know her own cognitive states directly by attending to them.

If we focus on the epistemic richness of the cognitive aspect of Mary's color knowledge, and grant that she is Highly Attuned, we can conceive of Mary learning all the physically reducible facts there are to know about the cognitive states of color perceivers at as fine-grained a level as the minimum distinguishable difference between perceptual experiences of colors. And it seems conceivable that Mary could learn to recognize these cognitive states in herself (or at least the cognitive states to which these more complex states reduce, since Mary has never been in the cognitive state one is in when one has a perceptual experience of color). So in knowing all the physical and physically reducible facts there are to know, Mary will know all the facts there are to know about human cognitive states and their relation to color stimuli. And if Mary is Highly Attuned, she will know her own occurrent cognitive states by directly attending to them. Highly Attuned Mary will therefore make the appropriate connection between her introspected cognitive state and her first perceptual experience of color, and thereby be able to immediately identify her first color experience.

4.2. Highly Attuned Mary and property dualism

The notion of a subject being Highly Attuned does not itself beg any questions against property dualism or the knowledge argument. Ordinary

¹²There are subjective and objective mdds – we can stipulate that she learns all the facts related to all the known objective and all reported subjective mdds in color perception without begging any questions against the anti-materialist.

¹³Those who hold a view that, as an empirical fact, any form of introspection will be accompanied by a phenomenal element or have some phenomenal component may still grant Mary this form of Highly Attuned introspection by stipulating that the phenomenal component of introspection is not essential to the knowledge acquired via these introspective acts.

subjects would not be able to achieve this Highly Attuned state in practice, but neither are ordinary subjects cognitively equipped to learn all the facts that, by hypothesis, Mary must learn in order to know all the physical facts there are to know about the perceptual experience of color. The fact that we do not conceive of normal subjects with such a heightened degree of awareness of their cognitive states and the ability to distinguish changes in themselves as fine-grained and quickly as being Highly Attuned would require says nothing about the conceivability of a subject being Highly Attuned.

Not only is the notion of being Highly Attuned compatible with the knowledge argument, it is compatible with the fundamental claims of property dualism as well. The property dualist's metaphysical separation of cognitive properties from phenomenal properties, underwriting the sharp divide between physical knowledge and phenomenal knowledge, actually encourages us to interpret Mary as being Highly Attuned. Mary's scientific omniscience will provide her with all the non-phenomenal information there is to have, and any facts about Mary's cognitive states that would be revealed through being Highly Attuned to those states would be non-phenomenal facts. Every aspect of the notion of being Highly Attuned is *entirely cognitive*, thus physically reducible; the states that are represented by the facts Mary would know in virtue of her being Highly Attuned are all non-phenomenal states, and there need not be any phenomenal aspect to such knowledge.

It is worth noting that Zombie Mary, Mary's phenomenally bereft physical duplicate, will differentiate between the same distinguishably different shades of red as her conscious twin, from which we can conclude that the difference between, for example, the perception of one shade of red (call it red-455) and the next distinguishably different shade of red (red-456) cannot be an essentially phenomenal difference. If this is correct, then the state that both Mary and Zombie Mary are in when they identify *that color experience* as one of red-455 must be a cognitive state, and this cognitive state must be present (or available) at the personal level. If these cognitive states were not present at the personal level there would be no reportable difference between perceiving two qualitatively distinct colors. If Mary knows all the physically reducible facts there are to know about all the possible cognitive states of color-perceivers and their relation to color stimuli, and can identify these states in herself, then Mary will be able to identify colors on sight in her very first perceptual experience of color. If this is correct, then suppressed premise 1.5 of the knowledge argument (when Mary sees color for the first time, she will not be able to identify colors on sight *solely in virtue of what she learns before her first experience of color*) is false; the support for the crucial

second premise (upon being cured, Mary learns something, viz., she learns what it's like to experience visual redness) has disappeared.

The paradox of phenomenal judgment reveals the epistemic richness of the cognitive aspect of perceptual experience, and one consequence of this epistemic richness is that regardless of what we initially assume to be the fundamental metaphysical facts regarding perceptual experience, a 'Mary' with the right kind of knowledge and training such that she could appropriately fulfill her role in a thought experiment designed to undermine materialism will be able to identify colors on sight (rather than being surprised to learn that red looks *like this*). The question we must now ask is whether there is any remaining motivation for the claim that Mary learns something new when she has her first color experience aside from the brute intuition that the physical facts of the world do not exhaust what there is to know about the world. The knowledge argument was supposed to support this intuition – it cannot also explicitly rely on this intuition to prove its conclusion without risking circularity. Mary's ability to identify color on sight does not on its own prove that materialism is true, but it does nullify the argumentative force of the knowledge argument. In addition, conceiving of a subject such as Highly Attuned Mary, with her color-identification abilities, would make materialism seem positively conceivable (and potentially ideally so) in the way the anti-zombie parity arguments require.

One could object that this argument is simply an instance of denying the antecedent: if Mary cannot identify color on sight, then there are irreducible phenomenal facts; however, this does not entail that if Mary can identify color on sight, there are no irreducible phenomenal facts. This would be clearly fallacious, and it is not the claim being made here. The claim here, rather, is that Highly Attuned Mary does not seem to learn anything new upon her first perceptual experience of color, so the burden is now on those who would deny the positive conceivability of materialism: present a case that shows Mary *must* learn something new even though she *seems* to already know everything there is to know about the perceptual experience of color. We have an adequate explanation for why Highly Attuned Mary identifies color on sight and learns no new facts: the epistemic richness of the cognitive aspect of her perceptual experience. We now need an argument to show why this 'seeming not to learn anything new' is inconclusive, and why this apparently conceivable materialist scenario is merely *prima facie* conceivable. All that is required to undermine the modal argument is to show that materialism is positively conceivable (and plausibly, ideally so); Highly Attuned Mary illustrates just the sort of scenario that would show materialism to be conceivable in the significant sense.

5. How the weakened knowledge argument helps destabilize the modal argument

Chalmers has claimed that the knowledge argument, and its ilk, should compellingly establish the failure of logical supervenience. The case of Highly Attuned Mary shows that the knowledge argument does *not* establish such a failure; indeed, Highly Attuned Mary gives us reason to believe that materialism is positively conceivable; that is, all truths are potentially knowable in virtue of having complete knowledge of the physical facts. With the conceivability of materialism thus established, an anti-zombie parity argument may be used to destabilize the modal argument.

For the anti-materialist modal argument to succeed, it requires (i) that the conceivability of property dualism be admissible as a premise in the modal argument for property dualism, (ii) that the conceivability of materialism not be allowed to play an analogous role in a modal argument for materialism, and, relatedly, (iii) that CP be a priori, and as such indisputable. In order to defend the *a prioricity* of CP, the anti-materialist cannot admit a case in which the application of CP entails a contradiction. So proponents of the anti-materialist modal argument must argue that materialism is in fact inconceivable. Agnosticism on this point will not suffice; if some form of metaphysical dualism is true of the actual world, materialism is necessarily false – false in every possible world.

It is now apparent how acknowledging the epistemic richness of the cognitive aspect of perceptual experience destabilizes both the knowledge argument and the modal argument. First, Highly Attuned Mary's ability to identify her perceptual experiences of color on sight illustrates a scenario that shows materialism to be positively conceivable (with plausible claim to being ideally positively conceivable). Second, if it is legitimate for the conceivability of property dualism to be used as a premise in an argument for the truth of property dualism because property dualism is (plausibly considered) ideally positively conceivable, then (by parity) it is legitimate for the conceivability of materialism to serve as a premise in an argument for the truth of materialism, since (as I hope to have shown) materialism may be plausibly considered ideally positively conceivable. Third, if the conceivability of both theories can support arguments for the truth of their respective view, both zombies and anti-zombies must be conceivable, and – if CP holds – not merely conceivable but possible, which generates a contradiction: metaphysically identical creatures will be, by hypothesis, both conscious and not conscious. Thus the problems the epistemic richness of the cognitive aspect of perceptual experience generates for the knowledge argument

destabilizes the modal argument as well. With the positive conceivability of materialism established, the anti-zombie parity argument now shows that CP can be used to generate a contradiction even in ideal cases, and CP can no longer serve as a premise in the zombie-conceivability argument without additional defense.

6. Conclusion

There are two main claims in the argument presented here. The first is that the knowledge argument fails as an anti-materialist argument – at best the result is neutral between metaphysical views, and at worst it pumps intuitions in favor of materialism, since the causal closure of the physical world guarantees that Highly Attuned Mary will be able to identify color on sight solely in virtue of her complete physically reducible knowledge. The second is that Highly Attuned Mary's ability to identify her perceptual experiences of color on sight illustrates a scenario in which materialism is (plausibly ideally) positively conceivable, which in turn undermines the anti-materialist modal argument. Given that we arrived at the conclusion that Highly Attuned Mary will identify color on sight merely by (1) accepting the causal closure of the physical world and (2) acknowledging the epistemic richness of the cognitive aspect of perceptual experience (which follows from the metaphysical division between the phenomenal and cognitive properties of the mind), one might try to develop the Highly Attuned Mary case into an argument that naturalized property dualism is a self-undermining theory in general (since both (1) and (2) are accepted by naturalized property dualists). Though this view is not being defended here, perhaps it should not be dismissed out of hand. The argument presented here has merely shown how focusing on the epistemic richness of the cognitive aspect of perceptual experience destabilizes the two most popular anti-materialist arguments: the knowledge argument and the modal argument. What this might imply for naturalized property dualism more generally is a topic to be taken up elsewhere.

Acknowledgements

I am grateful to David Chalmers, Daniel Dennett, Bryce Huebner, Bill Lycan, Luke Roelofs and anonymous referees for comments on earlier drafts of this paper. I have also benefited from advice from Mohan Matthen, Jennifer Nagel, Ram Neta, Diana Raffman, and Jonathan Rick.

Disclosure statement

No potential conflict of interest was reported by the author.

References

- Brown, R. 2010. "Deprioritizing the a Priori Arguments against Physicalism." *Journal of Consciousness Studies* 17 (3–4): 47–69.
- Chalmers, D. J. 1996. *The Conscious Mind*. New York: Oxford University Press.
- Chalmers, D. J. 1999. "Materialism and the Metaphysics of Modality." *Philosophy and Phenomenological Research* 59: 473–493.
- Chalmers, D. J. 2002. "Does Conceivability Entail Possibility?" In *Conceivability and Possibility*, edited by T. S. Gendler and J. Hawthorne, 145–200. Oxford: Clarendon Press.
- Chalmers, D. J. 2010. *The Character of Consciousness*. Oxford: Oxford University Press.
- Dennett, D. C. 1991. *Consciousness Explained*. Boston, MA: Little, Brown.
- Dennett, D. C. 2005. *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press/Bradford Books.
- Dennett, D. C. 2007. "What RoboMary Knows." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*, edited by T. Alter and S. Walter, 15–31. Oxford: Oxford University Press.
- Frankish, K. 2007. "The Anti-Zombie Argument." *The Philosophical Quarterly* 57 (229): 650–666.
- Graham, G., and T. Horgan. 2000. "Mary Mary Quite Contrary." *Philosophical Studies* 99: 59–87.
- Jackson, F. 1986. "What Mary Didn't Know." *The Journal of Philosophy* 83: 291–295.
- Kirk, R. 2005. *Zombies and Consciousness*. Oxford; New York: Oxford University Press.
- Kripke, S. 1972. "Naming and Necessity." In *Semantics of Natural Language*, edited by G. Harman and D. Davidson, 253–355. Dordrecht-Holland/Boston-USA: Reidel.
- Lycan, W. G. 2003. "Perspectival Representation and the Knowledge Argument." In *Consciousness: New Philosophical Essays*, edited by Q. Smith and A. Jokic, 384–395. Oxford: Oxford University Press.
- Marcus, E. 2004. "Why Zombies are Inconceivable." *Australasian Journal of Philosophy* 82 (3): 477–490.
- Marton, P. 1998. "Zombies vs. Materialists: The Battle for Conceivability." *Southwest Philosophy Review* 14: 131–138.
- McGeer, V. 2003. "The Trouble with Mary." *Pacific Philosophical Quarterly* 84 (4): 384–393.
- Sturgeon, S. 2000. *Matters of Mind: Consciousness, Reason and Nature*. London: Routledge.
- Yablo, S. 1993. "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research* 53 (1): 1–42.