The Knowledge Argument is Either Indefensible or Redundant

Dr Tom McClelland – University of Warwick

Introduction

Jackson's (1982) formulation of the knowledge argument (KA) has had an inestimable influence on the discussion of consciousness and the apparent problem it presents for physicalism. A common objection to KA is the 'ignorance objection'. According to this objection, our intuitions about Mary merely reflect our ignorance of physical facts that are integral to the explanation of phenomenal consciousness (e.g. Dennett 1991; Stoljar 2006). Armed with the insights of a future science, Mary would actually be able to deduce what it's like to see red. We only have the intuition that Mary would learn something new because we don't know the things she knows. Those sympathetic to KA have brushed away the ignorance objection on the grounds that even though we don't know what the future science of consciousness will reveal, we do know what *kind* of facts it will contain and know that this is the wrong kind of fact from which to deduce facts about phenomenal consciousness. In other words, they suggest that the insight that underwrites KA is 'future-proof': it is not the kind of insight that could be displaced by new scientific knowledge. I argue that this strategy for defending KA is dialectically problematic and reveals a fundamental limitation of the argument.

To future-proof KA the anti-physicalist would need to identify some feature credibly possessed by all physical facts – both known and unknown – and demonstrate that phenomenal facts cannot be deduced from facts possessing that feature.[1] The problem for advocates of KA is that if they can establish something about the nature of physical facts that precludes them from entailing phenomenal facts, there would be no need for the knowledge argument: we could rule out physicalism without having to trade intuitions about the epistemic situation of a subject with complete physical knowledge. So if KA were to be future-proofed, it would thereby be rendered redundant. But if, on the other hand, KA is not future-proofed then the ignorance objection stands and KA is indefensible. My master argument against KA thus runs as follows:

1) KA is defensible only if it can be future-proofed.

2) If KA can be future-proofed then it is redundant.

[1] Of course, if physicalism is true then there's a sense in which phenomenal facts are physical facts, so a more accurate characterisation would be that anti-physicalists must identify some feature credibly possessed by all *non-phenomenal* physical facts. Even if phenomenal facts are physical, they are not the facts that Mary is given in her black-and-white room, so it's appropriate to limit ourselves to physical facts of the kind that Mary would find in her textbooks.

3) Therefore KA is either indefensible or redundant.

In the first two sections I make the case for '1' and '2' respectively, considering various potential objections along the way. In the third section I conclude that KA is either indefensible or redundant and reflect on what this means for our understanding of consciousness and the case against physicalism.

1. KA Is Defensible Only if it Can Be Future-Proofed

1.1. The Ignorance Objection

If Mary were locked in her black-and-white room with full access to *today's* physics and neuroscience, it should be pretty uncontroversial that she would not find herself in a position to deduce what it's like to see red. But clearly this would have little bearing on the prospects of physicalism (all but the bravest physicalists would readily concede that we don't yet know the physical facts that entail what it's like to see red). KA rests on the claim that Mary would be unable to deduce the relevant phenomenal facts even if she had access to an *ideal* science that contained every physical fact that could be relevant to the perception of redness. This ideal science would contain many of the facts mentioned in today's science, but it would also contain other facts of which we are currently ignorant. Some of these facts may even be radically unlike anything contained in our best current theories. Given our ignorance of these facts, how can we trust our intuitions about Mary's epistemic situation? How can we rule out the hypothesis that knowledge of these as-yet-unknown physical facts would equip Mary to deduce what it's like to see red?

The ignorance objection can be bolstered by appeal to historical cases in which arguments analogous to KA appeared plausible but were shown by later scientific discoveries to be misguided. Stoljar (2006) highlights an argument put forward by C.D. Broad for the conclusion that chemical facts (i.e. facts about how different elements combine) are not entailed by non-chemical facts (i.e. facts about the nature of the constituent elements). Broad suggested that even with complete knowledge of oxygen, and complete knowledge of hydrogen, one would not be able to deduce that the two elements would combine. From this epistemic premise he drew the metaphysical conclusion that chemical properties are *emergent* properties irreducible to non-chemical properties. But Broad's conclusion was false. We now know that the quantum-mechanical properties of oxygen and hydrogen explain why they combine, so with complete knowledge of the elements one would be able to deduce how they would interact. Stoljar draws the following lessons from Broad's mistake:

> Just as the chemical argument was plausible to him, so the knowledge argument is plausible to us. Just as it is mistaken to follow the chemical argument to its conclusion, so it is mistaken to follow the knowledge argument to its conclusion. Finally, just as Broad was ignorant of a type of nonchemical truth relevant to the nature of chemistry, so, too, we are ignorant of a type of nonexperiential truth relevant to the nature of experience. (2006, p. 140)

Different advocates of the ignorance objection take different views on the depth of our ignorance. Some simply say that we are ignorant of the physical facts from which the phenomenal facts can be deduced (e.g. Dennett 1991). Some make the stronger claim that currently we do not even have the concepts needed to frame the relevant physical facts (e.g. Stoljar 2006). Others go further still and claim that we don't even have the psychological faculties needed to acquire the necessary concepts (e.g. McGinn 1989).[2] Different authors also take differing views on the content of our ignorance. Some say that the relevant unknown facts are neuroscientific (Dennett 1991), some appeal to our 'Russellian' ignorance of the intrinsic nature of matter (Strawson 1994; this volume) and others maintain a cautious neutrality on the character of the unknown facts (Stoljar 2006). For our purposes there is no need to adjudicate on these matters. There is a diverse family of positions revolving around a simple objection to KA: that Mary's knowledge of as-yet-unknown physical facts would equip her to deduce what it's like to see red.

How can advocates of KA respond to the ignorance objection? A concessive response might be to qualify KA so its conclusion is that *either* physicalism is false *or* that physicalism is true and we are ignorant of the physical facts responsible for phenomenal consciousness. This weakened formulation of KA would be unsatisfactory. The burden of proof is on anti-physicalists to show that physicalism is false, so an anti-physicalist argument that leaves open the second disjunct is no anti-physicalist argument at all.[3] If an argument gives us a choice between adopting anti-physicalism and adopting

[2] If we don't have the psychological faculties needed to acquire the concepts with which to frame the physical facts responsible for consciousness, then it could be argued that Mary would be unable to gain complete physical knowledge. If her psychological constitution is like ours, then these physical facts would be hidden from her. However, the point of the Mary thought experiment is that she is an ideal epistemic subject. If her possession of complete physical knowledge requires her to have psychological faculties quite unlike ours, then this can simply be built into the thought experiment.

[3] It is worth reflecting on where the burden of proof is here. Since our knowledge of the physical facts that might underwrite phenomenal consciousness is dramatically limited, it is tempting to say that we ought to be agnostic between physicalism and anti-physicalism. However, the general considerations in favour of physicalism (mainly those pertaining to parsimony and causal closure) give us defeasible evidence in its favour. As such, in the absence of defeaters it is reasonable to assume that as-yet-unknown physical facts will entail the phenomenal facts. KA thus effectively begs the question against physicalism by illicitly assuming that new physical knowledge won't yield an explanation of phenomenal consciousness. Given our ignorance we must acknowledge that when all the physical facts are in it *might* transpire that physicalism is false, but in order to

physicalism then, other things being equal, we ought to choose physicalism, so KA doesn't put physicalists under any real pressure.

In order to defend itself against the ignorance objection, KA cannot be so concessive. A robust defence of KA must give us reason to believe that the physical facts contained in an ideal science will be no more suited to entailing all the phenomenal facts than are the physical facts described by our current theories. If our intuitions about Mary's epistemic situation are based merely on the fact that today's science doesn't come close to enabling one to deduce what it's like to be in a certain brain state, then those intuitions should not be trusted. But if our intuitions are instead based on an appreciation of the nature of physical facts as such, then those intuitions may give us a genuine insight into Mary's epistemic situation. This would be a 'future-proof' insight – the kind of insight that doesn't risk being displaced by future scientific discoveries.

The above makes a strong case for the first premise of my master argument against KA. If the ignorance objection stands unchallenged, then we cannot reasonably trust our intuitions about the Mary thought experiment and KA should be dismissed as unsound. If, on the other hand, KA can be future-proofed then it survives the ignorance objection to fight again another day. We can thus conclude that *KA is defensible only if it can be future-proofed.* Before I make my case for the second premise of the master argument more needs to be said about what future-proofing KA would involve.

1.2. The Conditions of Future-Proofing

The best way to understand what future-proofing would involve is to look at how advocates of KA have defended themselves against the ignorance objection. One defence is that besides being unable to explain the phenomenal facts with our current scientific theories, we cannot even *imagine* a scientific theory that would explain the phenomenal facts (Chalmers 1996). This distinguishes phenomenal consciousness from other as-yet-unexplained natural phenomena. Marine biologists don't yet know how phytoplankton share an ecosystem without wiping each other out, but we can at least *imagine* what such an explanation would look like. Yet when we try to imagine a scientific theory of consciousness that would allow Mary to deduce what it's like to see red, we come up short. So we can conclude that even though Mary is blessed with knowledge of an ideal science, she would be unable to deduce the relevant phenomenal facts.

This line of defence against the ignorance objection is persuasively undermined by P.S. Churchland:

---

meet its burden of proof KA needs to provide positive non-question-begging reasons to conclude that the physical facts *won't* entail the phenomenal facts.

> Adding I cannot imagine explaining P merely adds a psychological fact about the speaker, from which again, nothing significant follows about the nature of the phenomenon in question. Whether we can or cannot imagine a phenomenon being explained in a certain way is a psychological fact about us, not an objective fact about the nature of the phenomenon itself… (1996 p. 407)[4]

To see why the appeal to imagination fails as a defence against the ignorance objection, consider again Broad's argument for emergentism. Broad might have insisted that he couldn't even *imagine* a non-chemical explanation of the chemical fact that oxygen combines with hydrogen. But all this would have revealed is a failure of imagination on Broad's part and not any insight into the nature of chemical truths. The ignorance objection is premised on the observation that ignorance can lead our intuitions astray. Appeals to what we can and cannot imagine won't help fend off this objection for the simple reason that our ignorance can also influence what we are capable of imagining.

The failure of this appeal to imagination suggests that advocates of KA need to say something more substantive about *why* it is implausible that future scientific discoveries will allow Mary to deduce the relevant phenomenal facts. Advocates of KA are unconvinced by the ignorance objection because they think physical facts are simply the *wrong kind* of fact from which to deduce phenomenal facts. As such, it doesn't matter that we're ignorant of many of the physical facts that Mary would know. Those unknown facts are of the same kind as the physical facts with which we are familiar, and no facts of that kind are suited to entailing phenomenal facts. Of course, the credibility of this line of thought depends on how the notion of 'wrong kind of fact' is unpacked. I suggest that advocates of KA must identify a 'future-proof feature' of physical facts such that:

> i) There is strong reason to believe all physical facts have feature F.
>
> ii) We know that phenomenal facts have feature G.
>
> iii) We know that *in principle* there can be no entailment from facts with feature F to facts with feature G.

Claims about what kind of facts will be disclosed by future science are inevitably difficult to justify, so requiring us to *know* that all physical facts have feature F may be too demanding. This is why the first condition is framed in terms of having *strong reason to believe* that all physical facts have this feature. This qualification isn't necessary for the second condition: advocates of KA ought to know

[4] Churchland's comments are directed against the 'Hard Problem' rather than KA. Against KA, she actually takes the stand that even though Mary *would* learn something new, this wouldn't be at odds with physicalism (1996, p. 403). Nevertheless, her comments reflect something of the spirit of the ignorance objection to KA.

what it is about phenomenal facts that precludes their entailment by physical facts.[5] KA concerns a specific phenomenal fact – the fact that seeing red has the particular phenomenology it has – but the argument is meant to yield an insight into the irreducibility of phenomenal facts across the board, hence the generality of the second condition. It is important that the third condition includes the 'in principle' clause. If one merely had a hunch that facts with feature F cannot entail facts with feature G, this would not be enough to fend off the ignorance objection. The objector could simply respond that since we are ignorant of some of the physical facts with feature F, our hunch could merely reflect our ignorance. So to avoid begging the question against the ignorance objection, we need a future-proof feature that we can be sure precludes entailment of phenomenal facts in principle. Anything based on the observation that all *known* facts with feature F fail to entail facts with feature G would be inadequate: a deeper insight is needed that justifies the conclusion that even *unknown* facts with feature F will fail to entail facts with feature G.

If advocates of KA can identify a future-proof feature, then the ignorance objection can be dismissed on the grounds that we have strong reason to believe that unknown physical facts would not help Mary deduce what it's like to see red. If such a feature cannot be identified, then it is very hard to see how the advocate of KA could rule out the hypothesis that Mary's knowledge of as-yet-unknown physical facts would allow her to deduce the relevant phenomenal facts from her black-and-white room. Put simply, KA can be future-proofed if and only if a candidate future-proof feature can be found that satisfies the three conditions above.

Although a number of such features have been proposed in the literature, there are two features that have attracted significant support: the *objectivity* of physical facts and the *structural* nature of physical facts. I will present the *prima facie* case for regarding each of these as future-proof features. I will not evaluate whether these proposed features stand up to scrutiny: my master argument against KA is neutral on whether KA can in fact be future-proofed, so it does not matter whether you find either of these candidate future-proof features convincing. My purpose here is simply to illustrate what a rebuttal of the ignorance objection would have to look like.

[5] One interesting possibility is that one might identify a feature F of physical facts and know that facts with that feature cannot entail the phenomenal facts without being able to pin down *what it is* about the phenomenal facts that makes them unsuited to being entailed by facts with feature F. In this scenario, our justification bottoms out with intuitions about the nature of the phenomenal and there is no need to identify any feature F of all phenomenal facts. I bracket this possibility for two reasons. First, the future-proof features identified in the literature are such that a property G of phenomenal facts *is* identified. Second, the defence of KA would certainly be stronger if a property G could be identified that explains why the phenomenal facts cannot be entailed by the physical facts, so in the first instance we ought to pursue future-proof features that satisfy condition 'ii'.

The first candidate future-proof feature of physical facts is *objectivity*. At a first pass, objective facts are those that are understandable from many points of view. In order to understand physical facts about the brain for example, it is not necessary for you to have had any specific kind of experience. The physical facts described in today's science are objective. Moreover, it is plausible that *all* physical facts are objective - Nagel describes the physical as '…a domain of objective facts *par excellence*…' (1974 p. 442). This suggests that objectivity satisfies the first condition. Subjective facts are those that are understandable only if one adopts a certain point of view (see Crane, this volume).[6] In order to understand what it's like to be a bat, for example, it is necessary for one to adopt the bat-ish point of view and have the kind of experiences that bats have (Nagel 1974). Phenomenal facts are subjective. The case of Mary might be regarded as illustrative of this: it is only by experiencing redness for herself that Mary can learn what it's like to see red. This indicates that the second condition is satisfied. It is also plausible that subjective facts cannot be deduced from objective facts. From objective facts we can deduce further objective facts, but we cannot deduce ourselves into an unfamiliar perspective – we cannot reason ourselves into having some new kind of experience. But subjective facts, by their very nature, can only be learned if we have the relevant kind of experience for ourselves, thus we cannot deduce the subjective facts from the objective facts. This indicates that objectivity satisfies the third condition. There is thus a *prima facie* case for objectivity being a future-proof feature of physical facts, as captured by the following three theses:

i_{obj}) There is strong reason to believe all physical facts are objective.

ii_{obj}) We know that phenomenal facts are subjective.

iii_{obj}) We know that in principle there can be no entailment from objective facts to subjective facts.

The second candidate future-proof feature is *being structural*. At a first pass, structural facts are those that pertain exclusively to spatial, temporal and causal relations between entities. Non-structural facts are those that involve anything over and above spatial, temporal and causal relations between entities.[7] It is plausible that current science describes only structural facts: neuroscience, for instance, describes nothing more than spatial, temporal and dynamic relations between neurons

[6] There are many senses of the term 'subjective'. Here I have used the understanding of the term most pertinent to KA, but other senses of the term may also be relevant to our assessment of physicalism. Indeed, I have suggested elsewhere (McClelland 2013) that phenomenal states are subjective in the sense that there is something it is like to be in those states for the subject, and that it is this understanding of subjectivity that captures the deeper problem for physicalists.

[7] How best to draw the structural/non-structural distinction is a matter of some controversy. For an insightful assessment of the options see Alter (2015). For those uncomfortable with my simple characterisation of the distinction, I would reiterate that it is not especially important to the argument of this paper how the distinction is understood. The characterisation offered is meant to be more illustrative than definitive.

and other brain cells at various levels. Moreover, it is plausible that future science will also describe only structural facts as scientific descriptions of entities are based only on how those entities interact with our senses and measuring instruments (or with other entities that in turn interact with our senses and measuring instruments).[8] This indicates that being structural satisfies the first condition. Phenomenal facts are plausibly non-structural. Phenomenal facts involve the instantiation of phenomenal qualities, such as phenomenal redness, and phenomenal qualities cannot be characterised in purely structural terms - they are intrinsic properties that transcend any purely structural characterisation.[9] This indicates that the second condition is satisfied. Finally, it is plausible that '…from structure and dynamics, one can infer only structure and dynamics.' (Chalmers 2002, p. 259) Though the structural facts of microphysics might entail the facts of biology or neuroscience, these are plausibly still structural facts, just on a different scale to those described by microphysics. This indicates that being structural satisfies the third condition.[10] There is thus a *prima facie* case for concluding that being structural is a future-proof feature satisfying all three conditions. Call the following three theses the Structural Theses:

i_{str}) There is strong reason to believe all physical facts are structural.

ii_{str}) We know that phenomenal facts are non-structural.

iii_{str}) We know that in principle there can be no entailment from structural facts to non-structural facts.

It will be useful for the remainder of the paper to have one candidate future-proof feature in mind as I develop my argument against KA. I will focus on being structural, rather being objective, for two reasons: one, I think there are serious objections to objectivity being a future-proof feature that do

[8] There are several different routes to this kind of conclusion: the Russellian route (Russell 1927) is driven by the claim that the causal nature of perception means that it can only tell us about the causal structure of percepts; the Ramseyan Humility route (Lewis 2009) is driven by the claim that the nature of theoretical predicates is such that they can only be used to give structural characterisations of phenomena; the Kantian Humility route (Langton 1998; Jackson 1998) is driven by the claim that the affective nature of knowledge cannot disclose the intrinsic nature of entities. These views and others overlap in various ways but also make a number of independent claims. Since I'm not concerned in this paper with evaluating the claim that scientific descriptions are inevitably structural, I will not put these issues to one side.

[9] This is not to say that phenomenal facts do not involve structure. Clearly, facts about my current phenomenology will include facts about the structure of qualities in my field of awareness. The point remains, however, that phenomenal facts are not purely structural. In other words, there is always a non-structural aspect to our phenomenal states.

[10] One might say that it's *analytic* that the third condition is satisfied given what it means for facts to be structural and what it means for them to be non-structural. If so, this would offer a particularly robust satisfaction of the third condition. Nevertheless, the condition itself needn't be formulated in such a way as to demand this.

not apply to the structural view;[11] two, the structural option is perhaps the more influential position in the literature today, with leading anti-physicalists such as Chalmers (2010) regarding it as the cornerstone of the case against physicalism.

That said, it is worth addressing one prominent objection to the Structural Theses. Some have suggested that although science can only describe structural physical facts, there are also *non*-structural physical facts beyond the reach of scientific description (e.g. Pereboom 2011). Although we have no way of knowing these non-structural facts, we can infer that they exist because structural facts must be grounded in non-structural facts. So underwriting facts about the causal dynamics of physical entities, there are facts about the intrinsic nature of those entities. If this line of thought is accurate then the first structural thesis - 'i_{str}' - is false. This would mean that advocates of the ignorance objection can propose that facts about phenomenal consciousness are deducible from physical facts that include these non-structural facts, and that it is our ignorance of these non-structural facts that leads our intuitions about Mary astray.

This 'Russellian' response to KA is worth taking very seriously, and I have defended a version of it elsewhere (McClelland 2013). However, it is controversial whether this objection to KA constitutes a vindication of physicalism. Some might hold that these non-structural facts are not properly described as *physical* facts. If it is constitutive of physical facts that they are the kind of fact that can be uncovered by scientific enquiry, then the non-structural facts in question would come out as non-physical. Furthermore, even if non-structural facts are deemed to qualify as physical facts, we're left with an unorthodox version of physicalism with which many physicalists would be unhappy. It would respect the letter of physicalism by avoiding positing non-physical facts, but would violate the spirit of physicalism by positing facts beyond the reach of scientific enquiry. In order to side-step these worries, I will add the proviso that if KA can be future-proofed by an appeal to the structural nature of physical facts, then this would rule out an *orthodox* physicalism but would not rule out *Russellian* physicalism.[12] Ruling out orthodox physicalism would of course be a dramatic result for KA, so there is no harm in bracketing the Russellian view for current purposes.

2. If KA Can Be Future-Proofed Then It Is Redundant

2.1. The Redundancy of KA

[11] The most serious objection is that the subjective/objective distinction is fundamentally an epistemological distinction rather than a metaphysical condition, which makes it hard to justify the claim that physical facts must be objective. The view that some physical facts are subjective is developed in detail by Howell (2013).

[12] Chalmers suggests that the conclusion of anti-physicalist arguments like KA ought to be that *either* physicalism is false *or* Russellianism is true (1996). Here I am in line with his assessment.

We have seen that in order to future-proof KA, one must identify a candidate future-proof feature of physical facts that satisfies the three conditions specified. The problem for KA is that if such a future-proof feature can be identified, then there is no need for KA. If we have good reason to believe that physical facts are not the kind of facts from which phenomenal facts can be deduced, then we have good reason to believe that physicalism is false. To see this, consider again the suggestion that *being structural* is a future-proof feature of physical facts. If the three Structural Theses stand up to scrutiny, then anti-physicalists can offer the following *structural argument against physicalism*:

1) There is strong reason to believe all physical facts are structural.

2) We know that phenomenal facts are non-structural.

3) We know that in principle there can be no entailment from structural facts to non-structural facts.

4) If there is no entailment from the physical facts to the phenomenal facts then physicalism is false.

5) Therefore there is strong reason to believe that physicalism is false.

Premises 1-3 are simply the three Structural Theses, so their truth is guaranteed if being structural satisfies the three conditions of being a future-proof feature. Advocates of KA are already committed to the truth of premise 4 as this is what enables them to move from an epistemic premise about Mary to a metaphysical conclusion about physicalism. So if a formulation of KA future-proofed by an appeal to the structural nature of physical facts is sound, then the structural argument against physicalism is also sound. And if the structural argument against physicalism is sound, then KA is redundant. A future-proofed formulation of KA will include all the premises of the structural argument above *plus further premises* about Mary and her epistemic situation. But such an argument would be surplus to requirements as the simpler structural argument would suffice to demonstrate the falsity of physicalism. This is not to say that the structural argument against physicalism is actually sound. It is just to say that *if* a formulation of KA reinforced by the Structure Theses is sound, *then* the structural argument against physicalism is sound and that KA would therefore be redundant.

The same applies for any candidate future-proof feature one might use to reinforce KA. If objectivity satisfies the relevant conditions, then one can offer an objectivity argument against physicalism that makes no reference to Mary and her epistemic situation. An analogy might help capture the dialectical situation here. Imagine you need help lifting a heavy piece of furniture, but the person who helps you is so strong that they can lift the furniture by themselves, rendering your efforts

entirely surplus to requirements. The advocate of KA is in an analogous position with their project of refuting physicalism. The ignorance objection shows that KA is only defensible if suitably reinforced, but any suitable reinforcement can do the job on its own, thus rendering KA redundant. This is my initial case for premise 2 of the master argument against KA. To develop this case further, I will consider three potential objections to the claim that if KA can be future-proofed then it is redundant.

2.2. Objections & Replies

The first objection is that even if the structural argument against physicalism (or its equivalent for any other candidate future-proof feature) constitutes a self-standing argument against physicalism, it doesn't follow that KA is redundant. The structural argument taken in isolation gives us some reason to reject physicalism, but the structural argument in tandem with KA gives us *more* reason to reject physicalism. So long as the case against physicalism is better with KA than without it, then it is wrong to say that KA is redundant.

The difficulty with this objection is that KA more likely *hinders* the case against physicalism than helps it. As a general rule, it is better to take on no more commitments than one needs to justify one's conclusion. If the anti-physicalist takes on the commitments not just of the structural argument but also of KA, then they leave themselves open to objections to those further commitments. The Mary scenario is a contentious thought experiment. It raises challenging issues regarding: what's involved in having complete physical knowledge; what's involved in having ideal reasoning skills; what exactly Mary learns on escaping her room; whether what Mary learns is a new fact; and so on. Physicalists will challenge the assumptions anti-physicalists make about these issues. One option for anti-physicalists is to defend those assumptions against physicalist attacks. But an alternative, more dialectically advisable, strategy is to cut these assumptions loose. The case against physicalism needn't be weighed down by these commitments. Even if they are commitments with which anti-physicalists are comfortable, it is ill-advised to present physicalists with such easy targets. Chalmers notes that '…many of the common responses to those thought experiments have no clear application as a response to the simple [structural] arguments…' (2010, p. xv) By dispensing with the Mary thought experiment the anti-physicalist can side-step those responses entirely and present a leaner, more defensible, case against physicalism.

The second objection is that the structural argument against physicalism is not self-standing as it relies on KA for its plausibility. Alter advocates a version of the structural argument against physicalism but thinks that the plausibility of this argument is inseparable from the plausibility of KA:

> I noted that the structure and dynamics argument's three main claims suggest a deductive argument for the epistemic gap. But that argument is not independent of the considerations typically used to establish the gap: intuitions about the Mary case, zombie cases, and other such thought experiments. (2015, p. 6)

If this is true, then KA is far from redundant. The structural argument against physicalism only works in tandem with KA because the Structural Theses only get their warrant from the Mary thought experiment. As Alter puts it, our intuitions about Mary are 'epistemically prior' to the relevant claims about structure (2015, p. 7).

I concede that the Mary scenario might be used to motivate the three Structural Theses that drive the structural argument against physicalism. To motivate the claim that all physical facts are structural you might observe that Mary – a subject with complete physical knowledge – lacks knowledge of any non-structural facts, indicating that there are no non-structural physical facts. To motivate the claim that phenomenal facts are non-structural you might observe that Mary learns a phenomenal fact on leaving her room, yet already knows all the structural facts, so must be learning a phenomenal fact that is non-structural. To motivate the claim that one cannot deduce non-structural facts from structural facts you might observe that Mary cannot deduce non-structural facts about phenomenal consciousness from her complete knowledge of the structural facts.

However, just because the Structural Theses *can* be motivated by appeal to Mary doesn't mean that they *need* to be, or even that they *ought* to be. The premises don't need to be motivated this way because each can be motivated without reference to a subject in Mary's epistemic position. The claim that all physical facts are structural might be motivated by the epistemological claim that we can only gain knowledge of physical entities via how they affect us, meaning we can describe only their dispositions to interact with us and other entities (see footnote 8). The claim that phenomenal facts are non-structural might be motivated by introspection of the qualities of one's phenomenal states. The claim that one cannot deduce non-structural facts from structural facts might be motivated *a priori* on the grounds that facts about spatial, temporal and causal relations can entail only further facts about spatial, temporal and causal relations.

The above indicates that the structural argument *could* be motivated without appeal to KA. I would go further and suggest that it *ought* to be motivated without appeal to KA. The point of introducing structure as a future-proof feature of physical facts is to fend off the ignorance objection KA. But if the relevant claims about structure are motivated with reference to KA, then the ignorance objection comes back to haunt the anti-physicalist. Whatever one's prior commitments are regarding physicalism, it is clear that Mary knows a wealth of physical facts of which we are deeply

ignorant. So how, from our position of ignorance, do we know that Mary's complete physical knowledge won't include non-structural facts? How, from our position of ignorance, do we know that Mary would learn a new non-structural fact on leaving her room? How, from our position of ignorance, do we know that Mary would be unable to deduce non-structural facts about phenomenal consciousness from her complete knowledge of structural physical facts? If we have a deeper justification for these claims, then it is this deeper justification that underwrites our understanding of the structural and not our intuitions about the Mary scenario. If we have no deeper justification for these claims, then they should not be given much credence as they might simply reflect our ignorance of the physical facts.

Overall, this presents Alter with a serious dilemma: if the three Structural Theses can only be motivated by KA, then both KA and the structural argument against physicalism fail. KA fails because our ignorance of the physical facts known by Mary renders our intuitions about Mary unreliable, and the structural argument against physicalism fails insofar as it depends on those intuitions about Mary. If, on the other hand, the Structural Theses can be motivated without reference to KA than the structural argument suffices to refute physicalism and KA is redundant. The Structural Theses would put conclusions about Mary's situation on a much surer footing, but those conclusions would not need to be given any role in the case against physicalism.

We now come to the third and final objection to my claim that if KA can be future-proofed then it is redundant. A critic might concede that the structural argument makes a self-standing case against physicalism whilst maintaining that the Mary thought experiment is nevertheless needed to make the structural argument *vivid*.[13] Even if the Mary scenario isn't integral to the case against physicalism, it is still integral to the exposition of that case. The structural argument is dry and abstract and the Mary scenario ameliorates these problems by presenting interlocutors with a concrete scenario: a scenario that tends to elicit strong intuitions. On this view, the Mary scenario is *dialectically* redundant in that a sound argument against physicalism can be provided without it, but it is not *rhetorically* redundant as it is a valuable and effective tool for making the case against physicalism vivid and for persuading interlocutors of its force. The Mary scenario serves to aid understanding, even if it doesn't serve to make the case against physicalism more defensible.

[13] Though Chalmers suggests that the structural argument is what really drives the case against physicalism, he does specifically note that thought experiments like the Mary scenario serve to make the case more vivid (2010, p. xv). In the same passage, Chalmers also suggests that thought experiments like the Mary scenario are '…a useful technical device for making the arguments more formal and more analysable.' (2010, p. xv) I'm not convinced that KA is any more formal than the structure argument against physicalism. As for being more analysable, KA might have the advantage of helping to make physicalist objections to anti-physicalism more clear, but this would be covered by the claim that KA helps make arguments more vivid.

Wartenburg offers some helpful reflections on the role of *illustrations* of philosophical ideas, citing Kant's vivid image of pure understanding as a land of truth surrounded by a stormy sea of illusion. Wartenburg rightly notes that 'Kant's imagery here, though quite graphic, does not make a philosophical contribution to his argument, even if it helps a reader understand its general thrust.' (2006, p. 21) The image is not redundant, but nor is it part of Kant's argument in the first critique. Something similar might be said of the way in which movies can make philosophical arguments vivid. One might explain the case for Cartesian skepticism to an undergraduate audience, and then present them with *The Matrix* as a vivid illustration of a skeptical hypothesis. Even though the movie might help students understand skepticism and convince them to take it seriously, the movie shouldn't be regarded as part of the argument for skepticism. The argument for skepticism is self-standing, and the movie merely serves to help communicate its force.

Perhaps the role of the Mary scenario can be understood by analogy with the role of Kant's metaphor, or of the lecturer's movie presentation. Mary is a vivid illustration of the epistemic gap between structural physical facts and non-structural phenomenal facts. It serves to aid our understanding of the structural argument against physicalism, but doesn't make a contribution to the merit of the argument itself. So even though the structural argument is a self-standing argument against physicalism, the Mary scenario is far from redundant.

I see two possible responses to this defence of KA. The first is to say that the Mary scenario isn't even needed to make the structural argument against physicalism vivid and persuasive. The thought experiment risks distracting interlocutors with unnecessary questions about the details of the scenario, and the case against physicalism is best made without reference to such thought experiments. I think this response would be too strong. There is no denying that the Mary scenario is a vivid thought experiment with the power to elicit strong intuitions. If I were trying to convince a neophyte of the case against physicalism, I wouldn't want to present the structural argument without also presenting the Mary thought experiment.

The second response is to concede, as I do, that the Mary scenario is far from redundant when it comes to communicating the case against physicalism. But the Mary scenario having illustrative value should not be confused with KA having dialectical force. To use the Mary scenario in this way is to borrow a thought experiment from KA whilst dispensing with the argument itself. KA is intended as an argument against physicalism, and has been treated as such for decades. Using the Mary scenario to illustrate a distinct argument against physicalism is far from a vindication of KA. It remains the case that KA itself is redundant, even if the thought-experiment that underwrites it is put to illustrative use elsewhere. The conclusion of the master argument is that KA – the argument against physicalism offered by Jackson – is either indefensible or redundant. The concession that the

Mary scenario – the thought-experiment used to drive KA – can still serve a rhetorical function is no concession at all.

I have considered three objections to the claim that if KA can be future-proofed then it is redundant. Along the way, I hope to have clarified why KA would be rendered redundant and what exactly its redundancy amounts to. I have focused closely on being structural as a future-proof feature. I should reiterate that there is no commitment here to structure actually being a future-proof feature: the claim is just that *if* it is such a feature then it would render KA redundant for the reasons discussed. I should also reiterate that parallel considerations will, I suggest, apply to any candidate future-proof features. If objectivity were presented as the future-proof feature, we could formulate an objectivity argument against physicalism that renders KA redundant for all the same reasons. And the same goes for any other future-proof feature that the anti-physicalist might present as a candidate.[14] Overall, we are in a position to conclude that the second premise of the master argument against KA stands up to scrutiny.

3. Therefore KA Is Either Indefensible Or Redundant

The ignorance objection shows that KA is defensible only if it can be future-proofed. The considerations above show that if it can be future-proofed then it is redundant. We can thus infer that KA is either indefensible or redundant. I think there are three lessons we can learn from this conclusion, two of them negative and the third positive.

The first lesson is that KA should be dispensed with as an argument against physicalism. If you are skeptical about the case against physicalism being future-proofed then you should dismiss KA as indefensible. If you are optimistic about the identification of a future-proof feature then you should dismiss KA as redundant. Either way, KA should not be taken seriously as an argument against physicalism. The thought experiment that drives KA can still serve a valuable illustrative role in discussion, but it cannot be taken as a self-standing case for rejecting physicalism. It is worth noting that my objections to KA do not rest on any presumption of physicalism. As with so many philosophical debates, the debate surrounding KA includes accusations of question-begging from both sides. Physicalist objections to KA are often too easily brushed away as betraying a presumption that physicalism is true, or a failure to take anti-physicalist intuitions seriously. No such

[14] It might be suggested that KA has a more intimate relationship with other candidate future-proof features. Perhaps the case for regarding objectivity as a future-proof feature depends essentially on the plausibility of KA, meaning that KA is not rendered redundant. Although this possibility oughtn't be dismissed out of hand, I think the burden of proof is on the critic to show that matters are any different for the other candidate future-proof features.

response is available to the argument I have offered. I have argued that even if a compelling case against physicalism can be developed, it remains the case that KA ought to be dispensed with. I have offered an argument that concerns only the dialectical structure of the case against physicalism, and not the relative merits of physicalist and anti-physicalist views of phenomenal consciousness.

The second lesson is that what goes for KA also goes for the conceivability argument (CA) and any other argument against physicalism based on comparable thought-experiments. The conceivability argument asks us to imagine a complete physical duplicate who lacks phenomenal consciousness. This means that all the physical facts that hold for us also hold for our duplicate. The ignorance objection presents a serious problem for CA. We are ignorant of many physical facts about ourselves, so when we try to conceive of a perfect physical duplicate, our attempts to do so are inevitably constrained by our ignorance. How can we rule out the hypothesis that if we had complete knowledge of the physical facts, then we would find zombie duplicates inconceivable? The only way is to identify some future-proof feature shared by all physical facts such that discovery of any facts with that feature could not alter the conceivability of zombies: so long as the physical facts about me and my duplicate have that feature, it will always be conceivable that my duplicate lacks phenomenal consciousness. As should now be clear, the difficulty with this response is that if it can be made to work, then CA would be rendered redundant. But if it cannot be made to work, then the ignorance objection stands and CA is indefensible. Thus CA is either indefensible or redundant. Perhaps there is an illustrative role left for the zombie thought-experiment to play, but dialectically speaking CA must be dispensed with.[15] The same goes for versions of CA that appeal to qualia inversion and the like rather than to zombies.

The third lesson is a positive recommendation that we ought to attend more closely to the question of whether physical facts have a future-proof feature that satisfies the three conditions specified. Consider the following passage from Chalmers about the structural argument against physicalism:

> There is a sense in which the argument here, which turns on simple issues about explanation, is more fundamental than conceivability arguments involving zombies, epistemological arguments involving Mary in her black-and-white room, and the like…It is sometimes supposed that nonreductive arguments turn essentially on these thought experiments, but this is just wrong. In fact…I suggest that the thought experiments turn essentially on points about structure and function... (2010, p. xv)

[15] Heikinheimo & Vaaja (2011) offer a well-developed argument for the redundancy of the conceivability argument along these lines. They also assert that similar considerations apply to KA (2011, p. 24), so the current paper can be read as a vindication of that assertion.

I agree with Chalmers that the common understanding of the case against physicalism gets things upside down, and that in order to make progress in this debate we must turn things the right way around. If we want to establish whether phenomenal consciousness presents a threat to physicalism, we ought to turn our attention to whether there is a plausible future-proof feature of physical facts. If anti-physicalists can find such a feature, they would have a case against physicalism that avoids many of the pitfalls that have dominated discussion of KA and CA. If physicalists can systematically rebut the candidate future-proof features, they would have a defence of physicalism that no version of KA or CA could overcome. For the anti-physicalists, their project is likely to involve refining their characterisations of the structural/non-structural divide or the objective/subjective divide. For the physicalists their project will likely involve casting doubt on the claim that all physical facts have the putative future-proof feature, or on the claim that facts with that feature cannot entail the phenomenal facts. This paper does not take a stand on which side of the debate has the better prospects, but it does hope to offer a clear recommendation on how best to make progress in that debate.

References

Alter, T. (2015) 'The Structure and Dynamics Argument Against Materialism', *Nous*, Online First: DOI 10.1111/nous.12134

Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: OUP

Chalmers, D. (2002) 'Consciousness and its place in nature' in Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings.* Oxford: OUP, pp.247-272

Chalmers, D. (2010) *The Character of Consciousness*. Oxford: OUP.

Churchland, P. S. (1996) 'The Hornswoggle Problem', *Journal of Consciousness Studies*, 3(5-6), 402-408

Crane, T. *this volume*

Dennett, D.C. (1991) *Consciousness Explained*. Boston MA: Little-Brown Heikinheimo & Vaaja (2011)

Heikinheimo, A. & Vaaja, T. (2013) 'The Redundancy of the Knowledge Argument in *The Conscious Mind*', *Journal of Consciousness Studies*, 20, No. 5–6, 2013, pp. 6–26

Howell, R. (2013) *Consciousness and the Limits of Objectivity*, Oxford: OUP

Jackson, F. (1982) 'Epiphenomenal Qualia', *Philosophical Quarterly,* 32, 127-36

Jackson, F. (1998) From Metaphysics to Ethics: A Defence of Conceptual Analysis. Oxford: Clarendon

Press

Langton, R. (1998) Kantian Humility: Our Ignorance of Things in Themselves. Oxford: Clarendon Press Lewis, D. (2009) 'Ramseyan Humility', in Nola & Braddon-Mitchell (eds.) Conceptual Analysis and Philosophical Naturalism. Cambridge MA: MIT Press, pp.203-222

McClelland, T. (2013) 'The Neo-Russellian Ignorance Hypothesis: A Hybrid Account of Phenomenal Consciousness', *Journal of Consciousness Studies*, Vol.20, No. 3-4, pp. 125-51

McGinn, C. (1989) 'Can we solve the mind-body problem?', *Mind*, 98, 349-66Nagel, T. (1974) 'What is it like to be a bat?', *Philosophical Review*, 83, 435-50

Pereboom, D. (2011) *Consciousness and the Prospects of Physicalism*. New York: OUP

Russell, B. (1927), The Analysis of Matter. London: George Allen & Unwin Ltd.

Stoljar, D. (2006) Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness. Oxford: OUP

Strawson, G. (1994) *Mental Reality*. Cambridge, MA: MIT Press

Strawson, G. *this volume*

Wartenberg, T.E. (2006) 'Beyond Mere Illustration: How Film Can Be Philosophy', *The Journal of Aesthetics and Art Criticism*, Vol. 64, No. 1, Special Issue: Thinking through Cinema: Film as Philosophy, pp.19-32