Philosopher's Compass 1/1 (2025): 22–41.

Rethinking The Hard Problem, Naturalism, and Idealism

Nathan Wagester[1]

[1]California Polytechnic State University, California, USA, nathanwagesterwork1@gmail.com

ABSTRACT

Contemporary discourse surrounding the hard problem of consciousness is generally considered a conflict between physicalism and dualism. In this paper, I argue for a third option which has largely been ignored by contemporary discourse by claiming that current conditions for a satisfactory answer to the hard problem are flawed. I identify two key informative conditions: the directionality and conceivability conditions. I then show how all current options fail to satisfy these conditions. I claim physicalism cannot satisfy the conceivability condition due to the Knowledge Argument. I consider three plausible objections to the knowledge argument: Lewis' ability analysis, the old fact/new concept view, and Dennett's dissolution of the problem. All of which I contend fail. Second, I claim that dualism will not work either, due to the reasons outlined in Karen Bennett's paper "Why I am Not a Dualist" (Bennett 2021). Finally, I contend that the motivations for the directionality condition are dubious, and thus, we have reason to doubt it. Particularly, I argue that the directionality condition is motivated by a specific interpretation of the causal closure of physics that we need not accept. Since we have reason to doubt the directionality condition, a wide range of largely ignored solutions to the hard problem remain open to us. One of these options is idealism. Finally, I give a rough outline of an idealist theory and explain why it can solve a revised hard problem.

KEYWORDS: Causal Closure, Idealism, Physicalism, Knowledge Argument, Dualism, Hard Problem

1 | INTRODUCTION

As stated by Chalmers in his now infamous paper "Facing Up to the Problem of Consciousness," the hard problem is the question of why it is that conscious experiences of a specific type can arise out of complex neural processing mechanisms and physical processes (Chalmers 2010). This problem appears especially hard because of qualia or the what-it's-like character of conscious states. For instance, the feeling of this particular shade of redness or the feeling of happiness appears to have a qualitative character that mere physical states just don't seem to have. Do atoms or electrons have a what it's like to be them? Intuitively, for most, the answer is no. Hence, the obvious question arises: how in the world could something qualitative arise from something non-qualitative?

Before we can go about answering this question, we should make clear the informative conditions

under which the problem is considered solved. I take it there are the two following conditions:

(A) **The Directionality Condition**: Any satisfactory explanation of consciousness ought to have the qualia of mental states be grounded in non-qualitative physical states.

(B) **The Conceivability Condition**: Any satisfactory explanation of consciousness ought to produce the result that consciousness is inconceivable without its explanans. (Where proposition P is conceivable for S just in case S can imagine a world in which P is true and P is inconceivable for S just in case S cannot imagine a world in which P is true.)

I take it that the hard problem is said to be solved if and only if conditions A and B are satisfied. The existence of the Directionality condition is evidenced in the very formulation of the question and the assumptions of most everyone involved in the current discourse. From Chalmers, David Lewis, Bennett, to Weisberg all frame the question as having the direction of physical-to-mental (Chalmers 2010, Lewis 2005, Bennet 2021, Weisberg). The mental must be grounded in the physical if we are to respect physical science. The existence of the conceivability condition is evidenced by the various conceivability arguments for dualism throughout the discourse. Gertler's phantom pain argument aims to show that qualia are ideally conceivable without physical stuff, whereas p-zombie arguments aim to show that physical stuff is ideally conceivable without inner mental life (Gertler 2013, Kirk 2019). Of course, these arguments rely on the premise that conceivability implies possibility, and hence, so does the conceivability condition. However, we will assume that this relationship between the ideal concept and possibility exists throughout the rest of this paper. Essentially, the conceivability condition posits that consciousness is not possible without its explanans since we are engaged in a reductive explanation that seeks to explain consciousness through some physical feature.

Now that we have stated the conditions under which the hard problem is said to be solved, we are now in a position to evaluate various theories which claim to be capable of overcoming the hard problem. I first turn to physicalism. Before we move forward, I would like to note that this paper will solely focus on what I take to be the most plausible form of physicalism: reductive physicalism. Hence, I will be ignoring both non-reductive physicalism and eliminativism. In regard to non-reductive physicalism, I think that the causal-exclusion problem as stated by Jaegwon Kim is fairly conclusive against the position. I will not rehearse his well-known argument here. However, I agree that such a view, which holds that mental properties are not reducible to physical properties but are explained by them, is bound to slide into reductive physicalism or dualism (Kim 1989, 1992). In regards to eliminativism, the position that denies there are any mental phenomena to be explained, I believe that one needs to only open their eyes to see that such a position is false.

2 | THE INITIAL ARGUMENT AGAINST PHYSICALISM

As we discussed earlier, there are two main opponents in the current discourse which claim to satisfy conditions A and B. These opponents are physicalism and dualism. This section will discuss physicalism and why it fails to satisfy condition B. Since it fails to satisfy condition B, it will not work as a solution to

the hard problem.

There are many conceivability arguments against physicalism. In this paper, I will focus on the Knowledge Argument. This argument shows that facts about consciousness are a genuine addition to the physical facts. Thereby showing that physical concepts in principle cannot satisfactorily account for phenomenal concepts. Condition B is thereby a non-starter. The general form this argument will take is as follows:

1. Facts about consciousness are a genuine addition to the physical facts.
2. If facts about consciousness are a genuine addition to the physical facts, then the physical facts which exhaust the possible physical explanations of consciousness are ideally conceivable without consciousness.
3. If the physical facts which exhaust the possible physical explanations of consciousness are ideally conceivable without consciousness, then physicalism is false.
4. Therefore, the physical facts which exhaust the possible physical explanations of consciousness are ideally conceivable without consciousness.
5. Therefore, physicalism is false.

The obvious point of contention in this argument will be premise one. Premise two trivially follows from premise one. If some type of fact presents a genuine addition to another type of fact, then the concept that represents the former type of fact will be a genuine addition to the latter. Hence, the latter will be conceivable without the former. It follows that if mental facts are genuine additions to physical facts, then physical facts are conceivable without conscious facts. Mental facts are ideally conceivable without its supposed explanans. Premise three trivially follows from physicalism. The general thesis of physicalism is that physical facts exhaust all of the facts. Hence, if it does not exhaust all the facts and physical facts are conceivable without conscious facts, physicalism is false. The point of contention will be premise one. It basically begs the question against the physicalist. It assumes what the physicalist seeks to deny: that physical facts cannot account for conscious facts.

Support for premise one can be found in Frank Jackson's Mary thought experiment which is commonly known as the Knowledge Argument. The thought experiment is as follows:

> **Mary's Room Thought Experiment**: Imagine an ingenious physicist named Mary who knows everything there is to know about the neurophysiology of the brain. Mary knows every physical fact relevant to consciousness. Now imagine that she lives in a completely black and white room and has not/cannot experience color in the room. Once Mary leaves the room and comes to experience red the first time, it appears that Mary learns something new. Namely, she now knows what it is like to experience color (Jackson 2005, Nida-Rümelin et. al. 2019).

There are many ways to object to the conclusion of the thought experiment that Mary has learned about some extra-physical facts. Here I take a look at three of the most potent objections: the ability analysis, the old fact/new concept view, and the imagination objection. I assume that if these objections fail, then the best objections against the Knowledge Argument fail and that physicalism fails to satisfy the conceivability condition.

2.1 | The Ability Analysis

In his paper "What Experience Teaches," Lewis looks for a way out of the knowledge argument. He settles upon the ability analysis. The ability analysis argues that what Mary gains when she sees red for the first time are various abilities to remember and imagine (Lewis 2005). She now has various memories of redness and can imagine various sorts of objects with a red character. For instance, if she has seen a horse before, she can now imagine what it would be like for a horse to be red. This analysis appears very plausible in light of the distinction between knowledge-that and knowledge-how. It is plausible that one can have know-how without knowing that. For instance, there are plenty of great mechanics who have little understanding of the theoretical posits of engineering as a science. People can know how to employ moral concepts without knowing moral theory and so forth. It is also plausible that one can know-that without knowing-how. For instance, one can know the principles of good bike riding but in fact be a terrible bike rider. Thus, what the proponent of the ability analysis says in the case of Mary is that Mary gains know-how instead of knowing-that.

I have two independent worries about this analysis. First, it is not clear to me that Mary has actually gained a new ability to remember and imagine. She always had the latent capacity to imagine red. It is not as if Mary has learned a skill like riding a bike. She simply saw red and now has accessed her latent capacities. This gets us into territory regarding the metaphysics of dispositional properties that would lead us too far astray. In any case, it does not get to the heart of the issue, since Mary clearly has gained some sort of know-how whether or not it is specifically an ability.

Second, I pose the question to the proponent of the ability analysis: what exactly is Mary's gain in know-how about? Of course, the proponent of the ability analysis will initially be dumb-struck with this question. The abilities are about imagining and remembering! Remembering and imagining what? They will say red.

In response, I would simply press on my questioning: well what exactly is red? Lewis, a functionalist, will probably respond with something like a brain state that occupies a certain functional role (Lewis 1983). This answer is, at the very least, extremely unintuitive. It does not seem like facts about the experience of the color 'red' are exhausted by the functional role a certain brain state plays. To see why, just imagine a discussion of the experience of artwork using this terminology. Can the beautiful color-scheme of artwork really be chalked up to functional brain-states? It doesn't seem so. Or perhaps Jeremy Fodor, computationalist, will say that redness can be chalked up to turing-style computations in the brain (Resorcla 2020). I again merely point to the characterization of artwork under this model. To me it seems that the ability analysis only makes sense by presupposing Lewis's spooky "phenomenal information." If we suppose that there is some what-it's-like phenomenal information, then the ability we have gained is to imagine phenomenal information.

One response many physicalists have had to this line of questioning is that Mary has gained a new ability to represent the world and to discriminate between certain phenomena utilizing a new concept which is only applicable to first-order experience. This is a plausible program that ought to be taken seriously. In the next section of this paper, I review the phenomenal concept strategy in detail. In line with my previous remarks on the failure of the ability analysis, I indicate that the phenomenal concepts

strategy is doomed to fail due to the fact that it lacks phenomenal information to provide content to its phenomenal concepts.

2.2 | The Phenomenal Concepts Strategy

I will now consider the phenomenal concepts strategy, or the old fact/new knowledge view. In her paper "Phenomenal Concepts and the Materialist Constraint," Joseph Levin surveys the phenomenal concepts strategy in detail. This view says that while Mary has not learned a new fact, she has gained a new concept. That is, she has gained a new way to represent, or know about, a certain fact about the world, in the case of seeing red for the first time (Levin, 2007).

Levin describes three avenues the physicalist might take for the old fact/new concept view to make sense: a semantic, epistemic, and demonstrative view. First, he notes that while we might think that Mary has epistemic access to all the physical facts prior to leaving the room, she has gained a new semantic primitive. Through viewing red the first time, Mary has gained a new way of representing the world that cannot be analyzed into the way she had of representing the world prior to her gaining this phenomenal concept. She is now able to fix a reference to a real physical entity, light-waves or some other entity, through the phenomenal concept of redness instead. She has gained a new mode of presenting the real physical world.

However, as Levin notes, this view yields an implausible semantic theory of phenomenal concepts for the physicalists. The obvious objection is that in order for her new phenomenal concept to fix a reference to the physical world it must do so by means of some properties If we assume that Mary had access to all of the physical properties which explained redness prior to her encountering redness, then it seems that the property by which she fixes a reference to the physical stuff cannot itself be physical. If the property by which the phenomenal concept fixes a reference to the physical stuff is itself physical, then Mary wouldn't have really gained a genuinely new concept. I take it that this objection works against the semantic strategy and will not consider it further (Levin 2007).

Now we can take a look at the epistemic analysis of phenomenal concepts. The proponents of this view hold that while Mary has not gained a semantic primitive per se, she has gained an epistemic primitive. When Mary leaves her room for the first time and observes the color 'red', she gains a sort of unmediated access to lightwaves, or a physical entity, which she did not have prior to leaving her room. She has gained a phenomenal concept in the sense that she has gained a judgment that lacks an evidentiary intermediary. The phenomenal concept of 'redness' is not semantically primitive, but it does possess a sort of epistemic primitiveness. Every time we encounter 'red' immediately, we can then use the phenomenal concept of red. Essentially, it allows us to differentiate types of concepts based on the level of epistemic access we have to that entity. For instance, the phenomenal concept of 'red' is distinct from the physical concept insofar as they are applied to the same entity but at different levels of epistemic access. In this story, we can then explain away the features of red entities by appealing to the mediate concepts of physics, such as light waves. The immediate phenomenal concepts are explained and reduced to mediated, physical concepts which can explain how the phenomenal concepts act in experience.

This seems plausible. For instance, if I merely learn through a textbook that Napoleon existed and

lost at the battle of Waterloo, this seems different than if I had met him in person and observed the battle of Waterloo first-hand. In the former, my knowledge of Napoleon is mediated through a textbook and a line of histories. In the latter, I directly observe Napoleon myself. Perhaps we might say that my concept of Napoleon in the former is of a different sort than the latter. In the first case, I apply my concept of 'Napoleon' to *mediate* evidence, and in the second, I apply my phenomenal concept to *immediate* encounters with Napoleon.

However, there is a major issue here: giving an epistemic primitiveness to phenomenal concepts itself does not explain why the link of phenomenal concepts to their physical explanations should appear so arbitrary, which is the very issue that motivated the phenomenal concepts strategy in the first place (Levin 2007). After all, red need not appear red given the structural and functional properties of light-wavelengths or our cognitive processing mechanisms. Simply stating that the representation of redness is more epistemically primitive, does not itself explain why redness is so drastically different from representation of light-waves.

I think with Levin that this points to a flaw in the epistemic phenomenal concepts strategy in general. Stating that phenomenal concepts are epistemically primitive merely names a property of phenomenal concepts. It does not explain what makes them epistemically primitive. In fact, we might even say that it yields an implausible semantics of phenomenal concepts. If they are epistemically primitive, if they are justified independently, or immediately, what is this immediate sort of justification grounded in? What explains why it is immediately justified? For me, the intuitive answer is that some additional representational content is gained such that the phenomena can now be represented immediately. However, if this is the case, the epistemic strategy inherits all of the problems of the semantic strategy. Then, we must admit that some non-physical properties fix our reference to redness (Levin 2007).

The final strategy Levin looks at is the demonstrative strategy. As Levin recounts, the demonstrative strategy is fairly straightforward. Essentially, this view claims that perhaps phenomenal concepts work like demonstratives. A subject might know a claim of the sort 'Jonathan wrote the book' but not know that 'This person I am speaking to right now wrote the book'. While these propositions express the same genuine fact, the knowledge that these two claims express the same fact, that these facts are identical, appears to be a genuine addition to a subject's knowledge, even though the subject has not learned a new fact (Levin 2007).

We can plausibly see how this applies to the Mary thought experiment. While she has not learned a new fact per se, she has learned that 'The physical concept of redness as expressed by our best physical theories of the world' is identical to the 'this color she is experiencing right now.'

While this approach seems promising, it falls prey to the same objections posed to the semantic and epistemic approaches. Typically knowledge of this sort revolves around access to some new representational content. Taking the Jonathan example, the most plausible reason I learned that the 'person I am speaking to right now wrote the book' is due to me gaining a new way of representing Jonathan. I have met Jonathan in person, I have seen Jonathan, listened and spoke to him, it is this information which allows me to identify him with the previous concept of the author of Jonathan. I have added new intensional content to the concept of Jonathan that allows me to refer to him in additional scenarios. So in the Mary case, it appears that we cannot explain the identification of the physical concept of red with Mary's

acquaintance with red unless Mary has gained some new representational content. However, if Mary has gained a genuine new way of representing red, then it appears she has to fix her reference to red in these additional scenarios to properties which themselves are not physical.

Perhaps phenomenal concepts like 'red' are *thin* concepts. That is, the meaning of redness is exhausted by simply 'this color which is perceived in x, y, z scenarios' (Levin 2007). We might even contend that the meaning of red is exhausted by its extension and it actually lacks intensional content. Thus while there are no phenomenal properties by which phenomenal concept fixes a reference to red, red is applied demonstratively to various instances of red objects in experience.

I must say that I find this account to yield a very implausible semantics of phenomenal concepts. Namely, it will suffer from objections of co-extension in the way that set-nominalism suffers from co-extension objections. We can imagine a scenario, where all red items are also blue items. In this scenario, we are forced to conclude that red items are actually identical to blue items(at least phenomenally)! An apparent absurdity. Redness is clearly phenomenally distinct from blue! I contend that this absurdity results from the fact that when we consider phenomenal concepts to be thin we fail to give a legitimate theory of how colors can be differentiated phenomenally. The thin account of phenomenal concepts simply cannot capture the representational difference between red and blue.

While Levin, does not believe that his rejection of the epistemic, semantic and demonstrative strategies to be conclusive against the phenomenal concept strategy, I think the fact that all three accounts fail on the grounds that they do not give an account of the phenomenal information which allows us to refer to its object, gives us reason to suspect that any physicalist account of a phenomenal concept in the Mary case is doomed to fail (Levin 2007). Afterall, the physicalist who takes up this strategy must assume that Mary has access to all of the physical information needed to refer to red prior to leaving her color-blind room. Hence, if she is to refer to red at all, it seems that she must do so by means of some non-physical information. But if this is so, we have demonstrated that physical concepts are unable to account for Mary's gain in knowledge.

2.3 | The Imagination Objection

I think it will be useful now to dive into Daniel Dennett's view on the Mary problem now that we have looked at the traditional physicalist solutions. Dennett believes that Mary would actually learn nothing new after leaving her colorless room. Contrary to the story given by both ability analysis and phenomenal concepts view, Dennett believes that Mary would have the capacity to imagine what red is like prior to leaving the colorless room, based on all of her physical knowledge. He contends he can dissolve the knowledge argument by working out exactly what the assumptions of the knowledge argument entail. If this plausible objection works, then perhaps our central contention that phenomenal information is required to make sense of the Mary-problem is misguided, since there is actually no problem at all.

Before going further into Dennett's attempted dissolution of the knowledge argument, I think it will be useful to give a brief overview of Dennett's view of the mind. Daniel Dennett is primarily a computationalist about the mind. At a very basic level, computationalists are functionalists who believe that the mind is the software to the brain's hardware and that the brain is an extremely complex computer

which possesses an array of syntactic devices that process information in such a way so as to complete certain tasks, or objectives. While Dennett does not believe in the legitimate existence of the software of the mind, he does think that it is a useful way to describe the material processes occurring. However, I will not be addressing the complex ins and outs of Dennett's view in this paper. What is of interest here is his attempted dissolution of the Mary problem.

Specifically, Dennett claims that philosophers have been misled in their intuitions regarding Mary's gain in knowledge, because they have not fully fleshed out the implications of assuming that Mary knows ALL physical facts pertaining to color vision and completely grasps every neuro-biological theory in relation to it. Dennett contends that once we truly flesh out the implications of the thought-experiment, we will see that Mary doesn't actually learn anything new at all once leaving her room. Here is Dennett's first iteration of is his revised Mary-thought experiment in the Epiphenomenalism section of "Consciousness Explained":

> **Revised Mary Room**:Imagine that one day Mary is allowed to leave her room and the people that captured her painted a banana blue, in the attempt to trick Mary into thinking she was seeing yellow when she was actually seeing blue. However, Mary, knowing all the relevant facts about what exactly she should be seeing in each situation, dumbfounds her captors when she responds: You tried to trick me, I know that banana is blue! (Dennett 1991)

This revised-Mary thought experiment is an attempt to show that Mary will not have gained any new knowledge upon her leaving her room when the thesis of physicalism is sufficiently fleshed out. To see why Mary would know that the banana is blue we merely need to think very deeply about what her knowing all of the relevant neurological facts about color vision would entail. If she knows ALL of the relevant facts about color vision, then she will know exactly what brain-state she will be in, and how she will behave, when she is introduced to certain frequencies of light waves. According to Dennett, indeed Mary will precisely be able to predict exactly what impression each stimulus will have on her nervous system. Since she has adequate knowledge of the dispositional properties of her nervous system, she will not have gained any new knowledge upon leaving her room. She already knows how a blue banana should look to her. That is, she is capable of imagining the blue banana prior to leaving her room

I think there is an obvious point of contention here: it appears prima facie implausible that, even given all of her knowledge of color vision, physics, and neuroscience, that she would be capable of imagining color prior to leaving her room, even if she somehow could devise a strategy to guess what she would see upon leaving the room. We can imagine that there is a radio wave technology in the room that physically blinds Mary while she is in the room. She has spent her entire life in this room, but was able to learn and understand all of the relevant physical facts regarding color vision via touch and audio. It appears that in this scenario, even given her intense study of the brain, her investigation of the entire anatomy, biological structures, brain states in response to stimuli, etcetera, that it would be incredibly implausible to assume that she would have the capacity to then imagine what it would be like to see anything at all. She cannot because she quite literally lacks that sort of sensory apparatus while she is in the room. Hence, upon leaving the room she gains more knowledge than merely knowledge about what it is like to see colors: but

she also gains knowledge on what it is like to imagine geometrical features visually, to conceive of space visually, to see movement, as well as color

Dennett here might say that we are assuming the falsity of his theory. He might say his theory is just that adequate knowledge of the mental state of color is knowledge of the dispositional properties of her optical nervous system. All the ability to see color amounts to is a certain dispositional property of her optical nervous system. Since Mary has complete knowledge of the dispositional properties of her optical nervous system, she therefore will be capable of imagining, or knowing what-it's-like, to see red-even if she is blind. Given her near omniscience and ability to reason through neurobiology and physics, she would also be capable of devising a strategy to guess exactly what she will see when she leaves the room. Even if this is counter-intuitive, it is not necessarily false.

In response, I would say that his theory then yields such an implausible result in our thought experiment that it needs to be revised. If someone lacks the capacity to sense or be aware of a certain sensory medium-such as vision, taste, touch, hearing, or smell- there is very little hope for them in imagining what it would be like in that sensory medium.

Dennett himself recognizes this original thought-experiment is unsatisfactory. Which is why in his paper "What RoboMary Knows" he deploys an additional thought experiment in which Mary is conceived as a computational machine that works within his computational model of consciousness:

> **Revised Mary, Robo-Mary and Robo-Mary 2.0**: Robo-Mary is a mach-2 robot which has near-omniscience regarding the scientific facts of color vision. However, Robo-Mary has not yet had her color vision turned on, nor does Robo-Mary have the capacity to program herself into having color vision due to the fact Robo-Mary cannot access the pixels which are required to represent colors in her algorithms. However, Robo-Mary wants to know what seeing color is like and is able to code images into various shades of gray. Robo-Mary only has access to color values of 1-1000. Based on the grey scales, textures, and various shapes of objects which are presented to Robo-Mary, she is able to determine what color they would be if she had color vision. If a tomato of a certain shape, lighting, and color is presented to her, Robo-Mary is able to determine that she would see red if she were able to see red. Given her knowledge of all the neurobiological, computational, and physical facts regarding color vision, Robo-Mary is able to program a model of herself which would be able to interpret and understand colors. This Robo-Mary 2.0 which Robo-Mary herself programs, is capable of seeing color vision and is capable of interpreting color values greater than 1000 within her computational machinery. When Robo-Mary has her color vision and pixels turned on, she is not in fact surprised by any encounters with new colors, including redness, for she has already predicted them through robo-mary 2.0 (Dennett 2007).

This thought-experiment is very compelling. If it works, it serves to completely dissolve the knowledge argument. The core thesis of this thought-experiment seems to be the following: An entity which is capable of predicting what colors they would perceive if they could perceive it, could then program

another entity which is capable of perceiving those colors, and then imagine perceiving it themselves-thereby knowing what it is like to perceive colors.

First it is not clear to me that Robo-Mary's programming of Robo-Mary 2.0 would then allow Robo-Mary to perceive color herself. To see why, imagine a color blind computer scientist who is capable of acutely perceiving various gray-scaled colors and assessing what sort of color they would normally be perceived as. Hence, if he perceives a tomato, he would know-based on the gray scale it is presented in-that it would normally be perceived as red. He is very intelligent on this matter, but clearly cannot see red. Say he programs a computer to recognize redness based on more than grayscale pixels. Clearly, or so it seems, that although his program is able to recognize colors, our programmer is never able to see these colors himself. Hence, while his model is able to perceive colors, he will never know what it is like to perceive colors himself. He is not even capable of imagining colors!

Dennett considers such an objection. In sum, Dennett effectively states that we are assuming that phenomenal information is required to imagine colors. We are assuming the very thesis which he denies (Dennett 2007). However, if Dennett wants to play ball in the philosopher's court of intuitions, then he must accept that our intuitions regarding my programmer's thought experiment do not line up with his theory. I further do not think that we are assuming that phenomenal information is required to imagine colors, rather we are merely assuming that a programmer will not necessarily be able to imagine what they program in the same way that a mathematician who grasps the complexities of Riemannian Geometry may not be able to represent it imagistically in our inherent Euclidean imagistic representation of space. We have to imagine Robo-Mary as not being able to grasp colors. Since Robo-Mary cannot grasp the qualitative character of colors, but is only able to program it, she cannot know fully what Robo-Mary 2.0 knows.

Even if we do not accept a Riemannian rejection of Dennett's thought-experiment, there is a further issue: it does not appear that Robo-Mary is capable of deducing redness from her knowledge of the relevant physical and neurobiological facts. Robo-Mary must program a Robot-Mary 2.0, and then, based on this updated version of itself and understanding of itself, know about redness . In a word, Robo-Mary was unable to adequately deduce from her knowledge of physical facts, neurobiology, and color vision, the seeing of redness. If this is the case, then Robo-Mary was not capable of grasping color from her knowledge of all these facts. She had to go the extra step of programming an additional entity from which she could use as a model, to then know what redness is like. It follows from this that knowledge of all the physical facts does not deductively lead to knowledge of all color facts. It therefore seems implausible that physicalism is true. For, if we know all of the physical facts, it should follow as a matter of course that we know all of the facts about color.

Dennet does briefly consider this objection as well. However, his response is unsatisfactory at best. Dennett claims that he does not see why Robo-Mary should be required to be able to deduce the phenomenal notions of color from all the physical and neurobiological facts (Dennett 2007). However, I contend that this seems to be a trivial consequence of physicalism. If physical facts explain all other facts, it would seem like we should be able to deduce what color is like from the basic physical entities. But, as this discussion has shown, it seems that we simply cannot!

One might think that this relies too heavily on a strong interpretation of the Principle of Sufficient

Reason. That is, this relies heavily on the principle that if we knew everything about some state of the world at t_1, then we would be able to deduce everything about the state of the world at t_2. However, science has shown that this need not be the case. For instance, quantum mechanics seems to give rise to probabilistic outcomes. We cannot deduce from some quantum entity being state X at t_1, that will necessarily be in state Y at t_2, since it might also be in state Z at t_2.

While this may be true for token entities, it is clear to me that it does not hold for types. Even if we cannot predict that some particle will necessarily be in state Z or Y at t_2 based on it being in state X at t_1, it is a fact about reductive metaphysical explanations, and identity claims in particular, that if an entity of Type B reduces to an entity of Type A, then if we knew all of the facts about Type A we ought to be able to deduce all the facts about Type B from Type A. For instance, an explanation of the cohesive and adhesive properties of water should be able to be deduced from the chemical facts of H_20. The way entities appear visually should follow as a matter of metaphysical necessity from basic functionalist explanations. If the functionalist or computationalist takes color to be a functional or computational state, the color-experience should necessarily be identical to that functional or computational state. It follows then that if we assume that Robo-Mary has access to all of the functional/computational brain states which allow her a complete understanding of the neuroscience of color-vision prior to her programming Robo-Mary 2.0, and she cannot deduce the color vision of Robo-Mary 2.0 from her programming, then she does not indeed satisfy the conceivability condition and Dennett's account fails.

I think this failure of Robo-Mary to deduce Robo-Mary 2.0's vision from her initial set of facts indicates a general failure in Dennett's account in the way that both the ability analysis and old fact/new concept analysis have failed. Robo-Mary cannot deduce Robo-Mary 2.0 because she lacks the concepts that would allow her to deduce Robo-Mary 2.0. If Robo-Mary had access to this representational or phenomenal content, then her knowledge vision would have followed as a matter of metaphysical necessity. However, her functional understanding of color vision just didn't cut it

2.4 | Physicalism Fails To Be a theory of what-its-like

As hinted at in previous sections, I think the failure of each strategy lays bare something elusive about consciousness. Namely, that consciousness has a what-it's-like character that cannot be cashed out in terms of structural or functional explanations. This what-its-like character of conscious states is precisely the phenomenal information that I indicate physicalists lacked in their explanation of Mary's gain in knowledge. I suggest that perhaps the ability analysis, phenomenal concepts strategy, and Dennett's dissolution all fail because Mary's gain in knowledge is merely a straightforward gain in her factive knowledge. Since she has gained knowledge about new facts, she must have learned some non-physical fact and fixed this reference to this non-physical fact via phenomenal information. The most intuitively plausible story of Mary's gain in knowledge is the correct one.

To illustrate this point in a different way, let us turn back to section 2.1 Here I briefly mentioned that what is so strange about using functional roles of brain states to talk about art is that talk of functional roles are not theories of what-it's-like. Art is all about transmitting a certain experience, a certain what-it's-like, to the viewer and the failure of physicalist theories of mind to account for artistic description indicates

that they fail to be theories of what-it's-like. As Chalmers discusses in "Facing up to the Hard Problem of Consciousness" all theories so far have failed to be theories of what-it's-like or experience (Chalmers 2005). They have altogether ignored the hard problem. They have ignored the what-it's-likeness that is constitutive of phenomenal information. For every physicalist theory of mind we can always meaningfully ask: well is that really conscious, is that really all what-it's-likeness is? This what-it's-like character of consciousness seems to elude structural and functional reductions. Again, functionalist and computationalist theorists sound so strange when describing art in terms of their fundamental ontology because they are not actually talking about what is important in art, its what-its-likeness. It would be a very strange reality indeed if the feeling of the sublime one gets from a certain piece of romantic poetry is really just brain states. Hence, I take it that any physicalist theory of mind fails to satisfy condition B at an a priori level. Consciousness is always ideally conceivable without its supposed explanans because consciousness has a what-it's-like character that eludes the sort of reduction the physicalist is engaged in. With this in mind, we will now tackle dualism.

3 | DUALISM

In light of the failure of physicalist theories to be adequate theories of what-it's-likeness, dualism seems like a plausible option. For our purposes, we will focus on the dualism relevant to contemporary discourse: Chalmer's naturalistic dualism. By Chalmers lights, there are qualia, or fundamental mental properties, which are governed by psycho-physical laws. These psycho-physical laws track certain correlations between mental states and physical states and posit a fundamental link between the two whereby the mental states supervene on certain aspects of physical states (Chalmers 2005).

Chalmer's theory seems to respect the directionality and conceivability condition. It respects the directionality condition because it maintains that mental states supervene on physical states. Consciousness is dependent on physical stuff in a non-trivial sense. Second, it trivially respects the conceivability condition since mental states trivially cannot be conceived without mental states. Consciousness, in this case, is its own explanation since it is treated as a fundamental entity.

One might think that dualism falls prey to a modified knowledge argument (Lewis 2005). Imagine that Mary knows all the relevant facts of psycho-physics pertaining to color vision. She knows exactly what mental states are correlated to which physical states. Yet, she lives in a wholly black and white room. Does Mary learn something upon leaving the black and white room and experiencing red for the first time?

I think this modified knowledge argument fails. It fails because there is no reason for the dualist to assume that Mary can know all the relevant facts of psychophysics without actually experiencing color. The contention of the dualist proponent of the knowledge-argument is precisely that Mary needs to experience a certain color to have the adequate factual knowledge of that mental state. Since Mary fails to have the experience, she fails to have adequate factual knowledge of that mental state.

In any case, there is still something independently wrong with dualism. Namely, it fails to be a satisfying answer to the directionality condition. In the dualist picture certain mental states are hooked up to

physical states by fundamental psycho-physical laws. One might rightly wonder how these physical states hookup to mental states. In the words of Karen Bennett "How can postulating bridging principles answer the hard problem. Postulating bridging principles does not in itself tell us *how* consciousness arises from the physical, it just tells us *that* it does. Calling them bridging principles or psychophysical laws does not do any explanatory work. It just names the connection" (Bennett 2021). Essentially, Chalmer's dualist fails to produce a satisfying answer to the directionality condition because it is a mere restatement of the directionality condition. It tells us that consciousness depends on the physical without telling us why or how consciousness depends on the physical.

Chalmers might respond to our objection by arguing that we are asking too much out of his theory. When one claims something is fundamental or primitive they need no further explanation. He needs no further explanation of his psycho-physical laws because he treats them as fundamental. They are the explanation!

I think this response fails. Certainly, we are entitled to treat some entities as fundamental. We are entitled to treat some primitive building relations, perhaps grounding, as fundamental. However, other strange or foreign entities which seem to beg explanation are not. For instance, A contemporary Humean might think that we need to explain modal properties in more fundamental terms. I take it that Chalmer's psycho-physical laws are one of these strange and foreign entities which just beg to be explained. Chalmers is not entitled to simply construct a new fundamental entity out of nowhere when we have no idea how it's supposed to work! For instance, Chalmer's proposes the principle of organizational invariance. This psycho-physical law states that two consciousnesses with the same functional organization will have qualitatively identical mental states. But why? What sort of connection does qualia have to functional organization? Chalmers can only respond that it is a primitive connection. Yet we have no idea what sort of connection this primitive connection is! Unlike say, mereological relations, where we have a clear idea of what sort of connection is going on, we have no clear sense of the sort of connection between physical and mental states. For instance, we have a sense of how a pepperoni is a part of a pepperoni pizza, but we have no sense of how the mental state of pain is related to neurochemical processes. Hence, we have strong reason to reject Chalmer's proposal as an ad hoc statement of the conditions under which consciousness is to be explained, without explaining consciousness. He simply states the directionality condition without actually giving a satisfying explanation of how it satisfies it.

4 | RETHINKING THE CONDITIONS

Every available attempt to satisfy the conceivability and directionality condition has failed. Physicalism fails to satisfy the conceivability condition, while dualism has not provided a satisfying answer to the directionality condition. There is little hope indicated by any traditional answer in the current discourse. Perhaps the failure to supply any sort of satisfying answer is due to something more fundamental than the answers themselves. Perhaps something has gone wrong with the conditions for a satisfying answer to the question itself. Something has gone wrong in either the conceivability or directionality condition

The conceivability condition is motivated by the fact that reductive explanations yield identity state-

ments and identity claims are thought to be necessary. For instance, Water is not just H_20, but water is necessarily H_20. My phone is not merely my phone, it is necessarily the specific phone that it is. Thus, consciousness is not just what it is reduced to, it is necessarily what it is reduced to. Since the conceivability condition is motivated by a non-controversial fact about reductive explanations, it seems trivial and not a target for re-thinking the conditions for a satisfying answer to the hard problem.

The motivation for the directionality condition is more dubious. The directionality condition is motivated by a commitment to naturalism. Specifically, an reductionist naturalism which states our metaphysics ought to solely respect the ontology posited by physics and that all facts necessarily reduce to physical facts. Thus, the mind must reduce to the physical stuff, because the hard sciences say that physical objects are the fundamental stuff. Since the mind is not a part of this list of fundamental physical objects, the mind must be explained by physical objects and not the other way around.

I take it that a naturalism of this sort is motivated by the presumed causal closure of physics. Formulations of the causal closure of physics typically take something of the following form: each physical cause only has physical causes, or each physical effect only has a physical cause (Jaegwon 1989, Spurrett and Papineau 1999, Shook 2011, Papineau 2023). This seems like a reasonable motivation for the directionality condition. If physical effects only have physical causes, and physics forms the fundamental scientific framework from which all other sciences are grounded, then it seems reasonable to conclude that all entities must at bottom be something physical.

I think there are two ways we can plausibly object to this sort of naturalism: we can either reject the reductionist causal program or we can provide a plausible alternative closure principle which respects its original motivation

First, as John Shook argued in his 2011 paper, a reductionist physicalist program faces innumerable difficulties and has several plausible alternatives. One of the primary difficulties faced by the reductionist is that it is in no way clear how we are supposed to reduce the claims of higher-order sciences-such as biology, economics, and geology-into physical terms. This might count as evidence for a perspectival realist view. Physics might be argued to be only one of many equally valid scientific frameworks (Shook 2011).

I would like to put this lively debate to the side. A unifying account of reality from which all the facts in reality can be grounded in more fundamental facts is certainly alluring, and I want to take part. I will first begin with an exposition on the motivation for the closure principle. I will then argue that we can sufficiently capture the key motivation for the closure principle while disputing the fundamental ontology of physics. Since we can respect the closure principle in this way, perhaps we have reason for doubting the directionality condition.

4.1 | Neutral Closer

The closure principle is motivated by the scientific fact that our best scientific theories about micro-physical and macro-physical entities such as atoms, electrons, stars, etcetera suggest that physical effects solely have physical causes. Principles such as the conservation of energy also indicate that the causal system of physics is closed. Nothing from outside of physics influences the causal structure of physical

systems. Due to the fact that the casual system of physics is closed, and all facts about macro-entities are supposed to reduce to micro-entities, then we must conclude that all the entities that exist are physical (Papineau 2023).

I want to note an important fact about physical closure, note that this is a metaphysical thesis about a certain ontological category posited by science. Specifically, we are assuming that physical entities, such as atoms and electrons, exist exactly as would be posited by an ideal science. These items have mass, charges, and exist within space-time.

However, I think we can plausibly object to this principle while retaining its original motivation. Specifically if we can propose a theory which can respect the causal structure posited by physics, expresses the same closure that physics does, and has all of the theoretical virtues typically associated with our best scientific theories, but does not share the same fundamental ontology, then this theory ought to be treated on a par with physicalism.

I think many will rightly be completely dumbstruck here. How in all the world could you possibly produce such a theory? The predictive validity of our current scientific theories relies precisely on the ontology that they posit! Of course, we have to accept the fundamental ontology of physics to accurately respect the causal structure of physics. I do not think this objection is as pressing as it may seem. For instance, we can imagine an idealism which treats the physical relations/properties as merely conceptual relations/properties. While physical entities actually reduce to conceptual entities, the structure posited by science is maintained. Instead of an electron existing, perhaps it is a point in experience, and instead of possessing an array of dispositional properties, perhaps it will possess an array of conceptual relations. It seems very plausible to me that we could produce such a 1-to-1 theory while retaining theoretical virtues such as parsimony. After all, isn't the very language in which physics expressed just mathematical and spatial models? Are these models not just conceptual models?

Perhaps many will remain unconvinced, which is why in the next section I have attempted to spell out the beginnings of an idealist theory in more detail. For the purposes of this section the important point is this: a theory which does all the work which physics can do and expresses closure ought to be treated at least on a par with physics. If we accept the possibility of neutral closure -respecting the causal structure of physics while not positing the entities exactly how sciences posit them-then we are in a position to see how the directionality condition can be doubted. Perhaps we can sufficiently explain, or reduce physical entities to mental entities and retain the causal structure of the physical world. If so, then we have undermined the directionality condition because this fundamental ontology is capable of doing all of the theoretical labor of the fundamental ontology posited by physicists. Furthermore, since physicalism has failed to satisfy the conceivability condition, idealism actually appears as the more plausible alternative.

5 | END-GAME: IDEALISM

I will now begin my exploration into one idealist framework. While idealism has remained relatively unpopular among most contemporary theorists, there have been some stirrings of idealism in the liter-

ature. For instance, Chalmers in "The Mind-Body Problem and Idealism" explores three idealist views and assesses their viability as solutions to the hard-problem of consciousness. Specifically, Chalmers takes a look at micro-idealism, macro-idealism, and cosmic-idealism. As the name suggests, micro-idealism holds that all facts are grounded in facts about micro-subjects, macro-idealism holds that all facts are grounded in facts about macro-subjects and cosmic idealism holds that all facts are grounded in facts about a single cosmic subject (Chalmers 2019).

There are various pluses and minuses to each of these three views. For our purposes, I will be sketching out a view that Chalmers all but dismisses. While Chalmers does think that an idealist position is viable, he focuses primarily on realist, non-phenomenalist idealism. Chalmers contends that idealism has gotten a bad rap in the literature precisely because it is usually associated with a sort of phenomenalism where facts about P are grounded in the appearance that P. Propositions are seen as claims about possible experience. When I make a claim that all apples are red, I am really saying that all apples in experience appear red. But, as Chalmers notes, there is an explanatory worry here. The phenomenalist has nothing to ground the stable structure of experience. While it is true that if I look at an apple from one angle, and then another, it will appear to me as the same apple. What exactly explains that? The realist has an obvious answer: the mind-independent physical world. However, the phenomenalist has nothing to appeal to beyond experience. (Chalmers 2019).

I do not think the problem is as potent as Chalmers thinks. While it is true that we cannot ground the coherent structure of experience by appealing to mere sensation alone, we can explain the coherent structure of experience by an appeal to concepts or higher-order thought processes. I think Chalmers is confusing sensation with experience. Chalmers is conceiving of experiences as points or moments which are connected together by some common thread. Specifically, I interpret him as contending that a phenomenalist idealism could not explain how an individual perception of an apple at time *t*, Region R in our visual field, Sensation(touch, audio, taste, etc) S, etc., could be related to that same perception of an apple at time t_2, Region R_2, and Sensation S_2, other than stating that the facts about the subjects perceiving of an apple are mere brute facts. Thus, Chalmers asks: what connects one experience to the next? What makes the structure of experience stable in a phenomenalist idealist picture? However, I do not think we need to accept this moment-to-moment view of experiences. Instead, we can view the experience more like a river that flows downstream. There aren't any points which need to be connected, since experience is one fluid process where one part of experience cannot be cleanly separated from the other. Instead, experience comes with its own structure. It comes with concepts which give the stream of experience its borders and and intentionality which gives it direction. For instance, it is the concept of apple which connects the variety of sensations which pertain to apples from moment to moment. While sensations may have this moment-to-moment, instantaneous structure, sensation in relation to concepts, experience, does not!

To give a very rough picture of how I think concepts can connect sensations and give experience 'borders'. We can take a look at the concept of 'chair'. Let us define the concept of 'chair' as any object which can be used for the purposes of sitting. We can then use this concept over time to identify an object, keep that identification of that object as a part of our continuous representation of that object over long periods of time, as well as to identify new particular objects as also being chairs. For instance, when a

subject encounters a couch composed of the collection of sensations $S_1 - S_n$, the subject can identify it as belonging to the concept 'chair', thereby "drawing a border" or giving identity conditions to $S_1 - S_n$ as opposed to all the other available sensations. I take it that a process of this sort grounds all experiences and is capable of grounding the stable structure of experience.

I will note that the sort of phenomenalist idealism I endorse here can be seen as a weak-phenomenalist thesis. I do think that metaphysical claims, or claims about certain universal scientific laws, ought to be seen as claims about concepts and the structural role they embody. However, this structural role is only relevant insofar as the claim itself is about how concepts constitute the ideal possible experience. Hence, facts about P are grounded in the appearance that P and the sort of mental structures that play into constituting the appearance of P.

With this response to Chalmers's threat to anti-realist idealism in mind, I would like to present my pet-theory. Our idealist theory will be defined with the following conjunction of theses:

1. **The Fundamentality of Subjects**: The fundamental entities in the world are thinkers, the particular objects in experience, and concepts.

2. **The fundamentality of Subjective Relations**: The fundamental building relations are intentional relations, 'felt' relations in experience, and conceptual relations.

Now, there is much here which begs to be explained and the heavy-duty work of spelling out idealism is outside of the scope of this paper. In any case, I think we can begin to see how the world will look like under an idealist picture of reality. We have both the entities and their building blocks. Thinkers in relation to objects of thought, mediated by the concept of the object, are going to be the fundamental entities in our idealist world. In fact, these three elements make the very world itself possible. The world is just the irreducible relation between subject and object. Structure is added to the world by the various fundamental subjective relations. We have the aboutness of the subject, the intentional relations, which make a variety of intentional experiences possible. We also have the sort of relations we feel in everyday experience. For instance, in the act of me typing on this keyboard right now is a relation that is 'felt' in experience. These sorts of relations are going to account for various activities that subjects engage in, in relation to the object of the subject. I am the subject, which is in a 'felt' relation to the keyboard. Finally, we have conceptual relations. Conceptual relations are relations between concepts. These are sort of relations we find in ideologies or in systems of ideas. The conceptual relations can be seen as more fundamental in that they non-reductively make sense of intentional and 'felt' relations. While I need a concept of myself and a concept of my keyboard, and a conceptual relation between the two, to make sense of the act of 'typing on my keyboard', the act in experience itself, the feel-y part, does not reduce to the conceptual relations.

With this theory in mind one may plausibly wonder how it addresses the hard problem. In the picture of the world cast by this theory the hard problem now becomes: how does the causal structure of the natural sciences arise out of 'felt', intentional and conceptual relations? The hard problem becomes inverted. Instead of thinking of the link as physical-to-mental we now think of the link mental-to-physical. I take it that physical relations can be reduced to conceptual relations in conjunction with 'felt' relations. For instance, a certain physical interaction between particle A and particle B, is now cast in light of a certain

'felt' relation mediated, or made sense of, by a certain conceptual relation. Say that A moves a certain distance. This 'moving' is cast as a 'felt' relation, a feel-y relation, in experience. The moving is made sense of by a certain structure. We have various mathematical equations that model the moving of A in space. These mathematical equations are just conceptual relations which add an extra layer of structure that allows the subject to make sense of, to understand, the feel-y relation. Thus, physical laws or Natural Laws, are actually just relations between certain sorts of concepts. Hence, the physical law that makes sense of the feel-y relation says that a certain kind A will act such and such under X and Y kinds of conditions. The relation in a natural law is a relation between kinds- a relation between concepts. The various feel-y relations and their objects are experiences, particulars, which are subsumed under certain regular laws-conceptual relations- which are described in the natural sciences. Hence the lawful regularity of A acting under conditions X and Y is the subsumption of A, X and Y under a certain set of concepts and the relation which holds between them. We learn of these conceptual relations, or natural laws, through the appearance of objects in experience.

We might note that this somewhat mirrors Armstrong's realist view in which natural laws are taken to be relations between universals. The inherent behavior of particular entities is externally determined and necessitated by universals. Of course, Armstrongian universals are immanent and are multiply instantiated at various points of space at once, and, importantly, Armstrong is a realist (Armstrong 1983). However, I think the important comparison here is this one: concepts provide a stable, plausible platform to ground experiential stability and explain natural laws.

One might wonder how we will be able to account for error in this model. In what sense can we be wrong if how things appear is the standard of natural law itself? This issue dissipates when we note that our perspective is not ideal and our appearances are not ideal. The natural laws, or conceptual relations, are grounded in an ideal appearance of the world. Perhaps, something like what appears to the divine or some kind of cosmic subject.

I take it that we now have a plausible account of the reduction of physical facts to mental facts. Physical relations are spelled out in terms of felt and conceptual relations. If we are still doubtful of the plausibility of the reduction of physical to mental, I point to Kant as a plausible way in which the physical can be spelled out in mental terms. For Kant, Newtonian mechanics is vindicated by the categories of the understanding in conjunction with the forms of intuition (Stang 2016). This is what I am getting at here with felt and conceptual relations

Before I finish, I would like to take a look at an objection to this new project. Namely, it would be an objection to this project using a reverse conceivability condition. The reverse conceivability condition is as follows:

> **Reverse Conceivability Condition**: Any satisfactory explanation of physical structure ought to produce the result that physical structure is ideally inconceivable without its explanans. One might argue that any idealist or pan-experientialist theory fails to satisfy the reverse conceivability condition. That is, it seems that we can conceive of a physical world without consciousness. If so, our new theories are doomed to fail.

This objection is potent and dangerous. If it works then it seems we are at a loss at how to solve the

hard problem. However, I think we can produce a response. I think it doesn't work because at any point of imagining a world without consciousness we are imagining the world perspectivally. We are taking a certain perspective towards that world. When I am imagining a world filled with p-zombies, I am thinking and representing that world in my mind. There is always a transcendental, constituting subject lurking in the background of an imagined world. I think that we can plausibly deny the possibility of imagining a world without consciousness, because it seems implausible that we can imagine a world without consciousness! Like we said earlier, under the idealist picture the thinker is a condition of possibility for any world. Hence, the imagination of a world without consciousness is a non-starter. There is always some thinker lurking in the back constituting it. In the case of a p-zombie world, the thinker constituting the perspective which characterizes that specific world is itself the subject who is thinking it. The illusion that a subjectless world is possible, arises from our ability to constitute and imagine the world ourselves.

6 | CONCLUSION

While I have not spelled out my idealist framework enough to definitively show that it can respect neutral closure, I think such a project ought to be explored due to the failure of physicalism and dualism to supply satisfactory answers to the hard problem of consciousness. I also think that there are several other candidate fundamental ontologies due to the expansive field of ideas now open to us. For instance, we can explore a Jamesian metaphysics of pure experience or further dive into the realist idealism proposed by Chalmers. What is important here is this: with a neutral closure condition we now have a new world of options through which we can better grasp the mind, the world, and situate our place between the two.

REFERENCES

Alter, T., & Walter, S. (Eds.). (2007). *What RoboMary knows* (pp. 15–31).

Armstrong, D. M. (1983). *What is a law of nature?*. Cambridge University Press. https://doi.org/10.1017/CBO9781316499030

Bennett, K. (2021). Why I am not a dualist. In *Oxford studies in philosophy of mind* (Vol. 1, pp. 208–231). https://doi.org/10.1093/oso/9780198845850.003.0008

Chalmers, D. J. (2010). Facing up to the problem of consciousness. In *The character of consciousness* (pp. 3–34). https://doi.org/10.1093/acprof:oso/9780195311105.003.0001

Chalmers, D. (2019). Idealism and the mind-body problem. In T. Alter & S. Walter (Eds.), *The Routledge handbook of panpsychism* (pp. 353–373). https://doi.org/10.4324/9781315717708-28

Dennett, D. (1991). *Consciousness explained*. Little, Brown.

Gertler, B., & Shoemaker, D. (2013). In defense of mind-body dualism. In M. Timmons (Ed.), *Norms, nature and knowledge* (pp. 108–125).

Jackson, F. (2005). What Mary didn't know. In *Philosophy of mind: Contemporary readings* (pp. 470–476). https://doi.org/10.4324/9780203987698-46

Kim, J. (1989). The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association*, *63*(3), 31–47. https://doi.org/10.2307/3130081

Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research*, *52*(1), 1–26. https://doi.org/10.2307/2107741

Kirk, R. (2019, March 19). Zombies. In *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/zombies/

Levine, J. (2007). Phenomenal concepts and the materialist constraint. In T. Alter & S. Walter (Eds.), *Philosophy of mind: Contemporary readings* (pp. 145–166).

Lewis, D. (1983). Mad pain and Martian pain. In *Philosophical papers* (Vol. I, pp. 122–132). https://doi.org/10.1093/0195032047.003.0009

Lewis, D. (2005). What experience teaches. In *Philosophy of mind: Contemporary readings* (pp. 479–502). https://doi.org/10.4324/9780203987698-48

Nida-Rümelin, M., & O Conaill, D. (2019, September 23). Qualia: The knowledge argument. In *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/qualia-knowledge/

Papineau, D. (2023). Naturalism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Fall 2023 Edition). https://plato.stanford.edu/archives/fall2023/entries/naturalism/

Rescorla, M. (2020, February 21). The computational theory of mind. In *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/computational-mind/

Shook, J. (2011). Varieties of twentieth-century American naturalism. *The Pluralist*, *6*(2), 1–17. https://doi.org/10.5406/pluralist.6.2.0001

Spurrett, D., & Papineau, D. (1999). A note on the completeness of 'physics'. *Analysis*, *59*, 25–29.

Stang, N. F. (2016, March 4). Kant's transcendental idealism. In *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/kant-transcendental-idealism/

Weisberg, J. (n.d.). The hard problem of consciousness. In *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/hard-con/