Taking Stock of Physicalism's Failure as a Theory of Mind

Noah K. Prentice

Department of Philosophy, Oregon State University

PHL 556: Minds, Brains, and Machines

Dr. Benjamin Stenberg

12 June 2024

Taking Stock of Physicalism's Failure as a Theory of Mind

This is a paper in the philosophy of mind. I will ultimately be arguing against particular theories of what the mind *is*, or what makes things *mental*. In the third section, I briefly discuss implications for the empirical sciences, but scientific progress is not my primary concern. Instead, this paper aims directly at the plausibility of theories of mind.

From the time of Plato to the time of Descartes, *dualism* dominated theories of mind. Roughly, dualism is the view that there are two kinds of substance: physical substance—which makes up rocks, wires, and brains—and mental substance, which makes up the mind. Now, dualism is not so dominant. It has largely been seen as implausible due to the mind-body problem, which is the problem of explaining or understanding how physical substance causally interacts with mental substance. This problem has led to a new dominant leader, physicalism, which consists of the claims that mental substance does not exist and that everything is physical. One of the main questions about theories of mind, then, is "Does physicalism fare any better than dualism?" In this paper, I argue it does not.

Before beginning the arguments, though, we should take a brief moment to discuss how we should evaluate physicalism's success. Physicalism is a theory of mind: its many varieties tell us what the mind *is*. When evaluating such theories, though, we must keep our epistemic vantage point in mind. As such, I evaluate the plausibility of physicalism as a philosophical theory from the human perspective. This means that true claims which fail to explain what they are trying to explain, fail to account for what they are trying to account for, or fail to justify what they are trying to justify, are considered failures.

Such is the methodology of the paper, but what claims will I be evaluating? Well, consider the difference between me and a rock. Among numerous non-mental, distinctively

physical differences, the main differences between me and a rock are (1) that I am *conscious* and capable of feelings like pain or the feeling of seeing the color blue, and the rock is not, and (2) that I am *thinking* and have beliefs, desires, understanding, hopes, etc., and the rock does not. These two differences are rightly taken to distinguish the mental from the non-mental, and so I will be evaluating physicalism's success on these particular topics. So, in the first section, I evaluate physicalism's success in accounting for consciousness and phenomenal experience, ultimately arguing that it fails to do so. In the second section, I evaluate physicalism's success in accounting for intentional states, arguing that it fails here, too. In the last section, then, I discuss the implications of my claims for science, knowledge, and causation.

1 Physicalism and Consciousness

Consciousness encompasses the phenomenal "feel" of experience. As Nagel (1974) put it, "fundamentally an organism has conscious mental states if and only if there is something that it is like to *be* that organism—something it is like *for* the organism" (p. 436). This includes what it is like to see the color red, the feeling of pain or orgasm, and even what it is like to be a bat. Importantly, consciousness is *subjective*. Direct observation of my feeling of pain is not like direct observation of a rock's falling, as only I can directly observe my feeling of pain while anyone with eyes can directly observe a rock's falling. In order for physicalism to succeed as a theory of mind, it must therefore account for consciousness despite its subjective nature.

1.1 Can Physicalism Account for Consciousness?

Due to consciousness's inherent subjectivity, it is very hard to study. It does not at first glance seem obvious how we are able to, for example, deliberately control our bodily movements, change our focus of attention from one thing to another, or tell the difference between a dog and a cat. Luckily, as Chalmers (1995) points out, these phenomena can be

explained in physical terms. This is why the questions of how such processes take place are what Chalmers calls the "easy problems" of consciousness: physicalism is able to answer the questions scientifically, at least in principle (p. 2). Chalmers says this is because

> they concern the explanation of cognitive abilities and functions. To explain a cognitive function, we need only specify a mechanism that can perform the function. The methods of cognitive science are well-suited for this sort of explanation, and so are well-suited to the easy problems of consciousness. (1995, p. 4)

So, when taking stock of physicalism's prospects, it is hard to understate its promise in answering these questions.

Unfortunately for physicalism, however, there is more we should ask of a theory of mind. After all, there is much more that seems to belong to the mental, and even to consciousness. For example, one of the questions above touch on what Chalmers calls the "hard problem" of consciousness: why it exists in the first place (1995, p. 3). Ideally, this would also have a physically grounded answer. If we could explain the existence of consciousness in the same way that we explain the existence of *life*, namely through scientific research in the relevant fields, then physicalism would also have promise in accounting for the existence of consciousness. This is not the case, though. As Chalmers writes, "even when we have explained the performance of all the cognitive and behavioral functions in the vicinity of experience […] there may still remain a further unanswered question: *Why is the performance of these functions accompanied by experience?*" (1995, p. 5). Chalmers's conclusion is strong. The point is not that we cannot understand the nature of consciousness by understanding the functions that it envelops or psychological role that it plays; instead, the point is that the mere existence of consciousness cannot be explained merely by explaining conscious psychological functions.[1]

[1] Chalmers (1996) also discusses this point and provides additional arguments for it.

Chalmers's point becomes clearer when we examine a similar point that physicalists make. The mind-body problem is posed by physicalists as a problem for dualism. The idea behind the mind-body problem is that, if mental and physical substances exist, it is not clear how they can causally interact. In particular, it seems that the creation of a physical thing—namely a human being—gives rise to the creation of a mental thing—namely a mind—but it is not clear how this could occur if these things have essentially different characters. It seems that it would be impossible for the dualist to explain how mental substance could arise from physical substance, and it seems that no set of physical facts would entail the presence of mental substance. The hard problem of consciousness is a very similar problem. The difference is that, for the hard problem of consciousness, the mysterious interaction is not interaction between mental and physical substances, but rather interaction between physical substances and subjective experiences. Nagel (1974) formulates this difficulty as follows:

> If physicalism is to be defended, the phenomenological features [of experience] must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. […] if the facts of experience—facts about what it is like *for* the experiencing organism—are accessible only from one point of view, then it is a mystery how the true character of experiences could be revealed in the physical operation of that organism. (pp. 437, 442)

In other words, the hard problem of consciousness is the problem of understanding how subjective experience could ever possibly arise from or be accounted for by purely objective phenomena: the more fine-grained and detailed our understanding of the brain becomes, the more objective facts we gain, but this puts us no closer to seeing how these objective facts would entail subjectivity. Just as no set of physical facts would entail the existence of mental substance if dualism were true, it seems that no set of objective facts would entail the existence of subjective experience if physicalism were true.

Upon recognizing this difficulty, the physicalist may wonder just how detrimental the hard problem is for their theory. As already established, consciousness is one of the two mental differences between people and rocks. Does the hard problem ruin all of physicalism's hope in accounting for this difference? On the one hand, the answer is obviously "no." After all, the easy problems are easy. As long as science progresses properly, questions of mental functions can seemingly be explained very well in physical terms. On the other hand, there is clearly *something* in the hard problem that constitutes a failure for physicalism. Just how widespread is the failure? And is it enough to reject physicalism in favor of dualism? This is the main question I will attempt to answer.

1.2 The Extent of Physicalist Failure in Accounting for Consciousness

I claim that physicalism's failure is almost total, and that we *are* justified in rejecting physicalism as a theory of mind. First, though, let us briefly discuss what we require from a good theory of consciousness. What do we require of theories in general? Well, consider theories of *gravity.* Theories of gravity attempt, first and foremost, to *account* for gravity. This involves making a justified claim that gravity arises from some other physical phenomenon. Newtonian physics does so by positing the existence of entities called *forces*, one of which, *gravitational force*, occurs between objects according to a property called their *mass*. Relativity denies that such a gravitational force exists; instead, it accounts for gravity in terms of the warping of space-time. Still, both theories account for gravity. If a theory fails to account for gravity, we reject it as a gravitational theory. If the existence of consciousness is assumed to be physical, we should expect physicalist theories to be able to account for consciousness in the same way that we have physicalist accounts for gravity. The hard problem of consciousness complicates this, though: if physicalism cannot *explain* consciousness, how can it *account* for consciousness? After all, these

notions are directly related. Any physicalist account of consciousness would involve a justified claim to the effect of "consciousness arises from physical phenomena X, Y, and Z," which would immediately provide an explanation of the existence of consciousness; since Chalmers shows that such an explanation cannot exist, neither can a physicalist account of consciousness.

I should re-emphasize that I am not arguing that all claims of the form "consciousness arises from physical phenomena X, Y, and Z" are *false*. I am merely arguing that no such claims can be justified. So, some such claims might be true, but we lack and can never gain the evidence required to make them. It is in this sense that physicalism has unavoidable limitations as a theory of consciousness. Since consciousness is one of the two most fundamentally distinguishing features of the mental, physicalism has unavoidable limitations as a theory of mind.

2 Physicalism and Intentionality

Due to the hard problem of consciousness, physicalism fails to account for consciousness. Since the mind is distinguished by consciousness and intentionality, this means physicalism is, at most, half of a proper theory of mind. This has given us a partial answer to our main question, which was "How widespread is physicalism's failure to solve the hard problem of consciousness, and does it warrant a return to dualism?" To complete answering the question, we have to assess physicalism's prospects in the other fundamentally distinguishing feature of the mind: intentionality.

2.1 What Intentionality Is

Many of our mental states are *about* things: my belief that it is raining seems to be about the fact that it is raining (or about the present state of affairs of it raining). This feature of our mental states is called *intentionality*. As a rough definition of intentionality, Searle (1983) provides the following: "Intentionality is that property of many mental states and events by

which they are directed at or about or of objects or states of affairs in the world" (p. 1). Of course, mental states are not the only things with intentionality: the word "rock" refers directly to rocks, a painting might be of a ship, and so on. The intentionality of these is not obviously related to the study of mind, though, and so we can put such cases aside for now and focus on accounting for the intentionality of our mental states.

Before assessing different accounts of intentionality, we should be clear on what exactly intentionality is. Searle and others admit the definition of intentionality as the "aboutness" or "directedness" of our mental states is preliminary or crude (1983, p. 1). The question is how to home in on this notion. Searle proposes the condition that intentionality is the feature of mental states that leads to questions like "What is [the mental state] about? What is [the mental state] of? What is it [a mental state] that?" (p. 2). Mendelovici (2018) makes a similar move, defining intentionality as "[that] feature, whatever it is, that we at least sometimes notice in ourselves and are tempted to describe using representational terms like 'aboutness' and 'directedness'" (p. 5). Mendelovici also fixes reference on intentionality by pointing to *paradigm cases*, examples of intentionality where the temptation to attribute an "aboutness" or "directedness" most obviously arises, such as in my perception of the color sky-blue, or my belief about the rain outside (p. 6).

2.2 The Phenomenality of Some Intentional States

With our reference to intentionality fixed, we can observe how at least *some* intentional mental states are fundamentally phenomenal. Take, for instance, desire. When I desire sunshine (or desire that it be sunny outside), I mean that I *feel* a particular way about the state of affairs in which it is sunny outside. Furthermore, this feeling is the kind of feeling where there is *something it is like* to feel it: there is something it is like to desire sunshine, and this is a fundamental component of what a desire is. If someone said they desired sunshine but claimed

that they didn't feel any particular way in their state of desire, we would say that they are mistaken.

Let us take a moment to consider what impact this has on physicalism's ability to account for intentional mental states. As I argued in section 1, the hard problem of consciousness constitutes a failure for physicalism to account for phenomenal experience. And, since desires fundamentally consist of a feeling of the "what it's like" kind, desires are fundamentally composed of an aspect of phenomenal experience (among other things). So, the hard problem of consciousness constitutes a failure for physicalism to account for the phenomenal component of desire and hence of desire itself. For the sake of clarity, let us look to an analogy. Suppose I come up to you and say, "I know what a knife is made of, and I can tell you how to make it." If, upon interrogation, I revealed that I could *not* tell you what the blade is, or how it was made, you would rightfully say that my understanding of knives themselves is incomplete—any account of a knife I were to give you that failed to account for its blade would be seriously lacking, as the blade is a fundamental component of a knife. So it goes for physicalist accounts of desire: since they cannot account for the phenomenal, experiential component of a desire, and since the phenomenal, experiential component of a desire is *fundamental* to what a desire is, they cannot account for a desire itself.

One might object here that I am conflating two aspects of a desire: there are questions of what a desire *is*, metaphysically speaking, but this might be different than questions of what *it is like* to have a desire (or to be in the state of having a desire). If we make this distinction, then a functionalist description of desire might seem plausible: to desire sunshine is just to be in a state that is likely to be caused by long, overcast winters and is likely to cause the system to be outside on a sunny day. Then, on top of the desire itself, there is something it is like to have the desire,

but this experience is not fundamental to the desire as I have claimed above. So, physicalist functionalism can account for desire all right, and it is only the phenomenal experience which suffers. But this is too far. After all, trees seem to possess functional states like the one given above, but nobody would claim that trees *literally* desire sunshine. Instead, such attributions are mere cases of anthropomorphizing or metaphor. Since we want to know what it is for a system to *literally* desire sunshine, it seems that an experiential component is necessary just as it is for pain or joy. Since physicalism cannot account for this experiential component, it cannot account for desire.

Desire is not a uniquely problematic intentional mental state for physicalism, either. Just as I might desire sunshine, so too can I hope that it be sunny tomorrow or fear the possibility of it raining tomorrow. All of these mental states are intentional, and yet phenomenal experience is one of their essential ingredients. So, if my arguments above are correct, there are in fact quite a few intentional mental states which physicalism fails to account for, making it even more limited than we concluded in section 1. But perhaps physicalism can at least account for the *property* of being intentional, even if it cannot account for all of the mental states which instantiate the property. This would be the least the physicalist could hope for; as I will argue now, however, they are not guaranteed even that.

2.3 How Intentionality Arises

Why are intentional things intentional? Take words, for instance. Why is a word intentional? The most obvious answer is that they *derive* their intentionality from thoughts, concepts, or perceptions: my concept of rain is about rain, and I use the word "rain" to refer to the thing that my concept is about. So, maybe the intentionality of the word "rain" comes from the intentionality of my thoughts about rain. This seems reasonable, but even if some intentional

things derive their intentionality from other things, there must be something that brings intentionality into the world for the first time. Intentionality that is *not* derived in this way is called *original*. The question for theories of mind, then, is how original intentionality arises.

2.3.1 *Tracking Theories*

One option might be that original intentionality arises from the intentional thing *tracking* or being caused by something else in the world. Here, tracking is taken to be anything including "detecting, carrying information about or having the function of carrying information about, or otherwise appropriately corresponding to items in the environment" (Mendelovici, 2018, pp. 33-34). This theory claims that, for instance, my thoughts about rain are intentional because they track or are caused by something in the world, namely rain. In this way, tracking relations not only tell us how our intentional mental states arise, but also what they *represent*. So, the tracking theory makes predictions about the content, or the meaning, of our intentional states.

Mendelovici (2018) argues that the tracking theory does not work. In particular, some of the tracking theory's predictions about the content of one of the paradigm cases of intentionality turn out to be false. The paradigm case in question is perceptual color representation, such as a visual state representing a sky-blue mug. When my visual state represents a sky-blue mug, the tracking theory predicts that the content of my representation is something like a surface reflectance profile, which is a disposition to "reflect, transmit, or emit such-and-such proportions of such-and-such wavelengths of light" (Mendelovici, p. 39). Other physical features of the mug might be candidates for the property being tracked, but Mendelovici's argument is the same: when we introspect on the content of our representation of a sky-blue mug, we see that the content does not contain surface reflectance or any other physical property. Instead, as Mendelovici claims, our perceptual color representations represent color properties that are

qualitative, simple, primitive, and non-relational (p. 42). This kind of color representation is what Chalmers and Mendelovici call edenic colors (Mendelovici, p. 42). This means that the predictions which tracking theories make about the content of our originally intentional mental states are, at least sometimes, false. So, original intentionality does not arise from tracking relations.

2.3.2 *Functional Role Theories*

Another theory of original intentionality is that original intentional states arise from mental representations' functional roles, that is, the causal and functional relations which our mental representations have to other things. The scope of these functional relations separates two kinds of functional role theory: the *short-arm* functional role theories restrict the relevant functional roles to those relating our representations to other representations or internal items, while the *long-arm* functional role theories take the relevant functional roles to include both internal and external items. On page 72, Mendelovici (2018) uses the example of the concept 'bachelor' to illustrate how the content of the concept 'bachelor' is allegedly fixed by its functional role:

(B1) From judging O IS A BACHELOR, one is likely to judge O IS A MAN.

(B2) From judging O IS A BACHELOR, one is likely to judge O IS UNMARRIED.

(B3) From judging O IS A MAN and O IS UNMARRIED, one is likely to judge O IS A BACHELOR.

For the short-arm functional role theorist, relations like these exhaust the defining functional relations for the concept 'bachelor'. For the long-arm functional role theorist, these relations may be accompanied by tracking relations or other relations the concept 'bachelor' has to the external world.

Mendelovici (2018) also argues against functional role theories. One problem she identifies for short-arm functional role theories is that it is not clear why the internal functional relations that define 'bachelor' would give rise to intentionality (pp. 72-73). The central claim of the short-arm functional role theory is that there is some kind of isomorphism—some kind of structural or functional similarity—between the relational network of our representations and the relational network of contents, and that this isomorphism gives the representations their associated content. But, as BonJour (1998) argues, it is not clear why the existence of an isomorphism between representations and contents should make the representations represent the contents (p. 177). Just because my representation of a bachelor and its content happen to sit in the same place in their respective networks, there is no reason why there should be a connection between the networks themselves.

Long-arm functional role theories might avoid this problem by invoking relations between representations and the external world: if one representation 'grabs' its content by its relation to the external world, then this content may be passed around to other representations by virtue of the functional role it plays. The problem, as Mendelovici (2018) argues, is that the long-arm function role theorist must specify the relation representations have with the external world, and it seems like all of the candidates have serious issues. For instance, if the long-arm functional role theorist takes representations to be related to the external world through tracking or causal relations, then it inherits the tracking theory's mismatch problems described above. Alternatives might include behavioral or dispositional relations directed towards external items, but these still run into mismatch problems with perceptual color representations: perceptual color representations have the content of edenic colors, but our behavior or behavioral dispositions are not directed to edenic colors. So, it seems that intentionality does not arise from functional roles.

2.3.3 *The Phenomenal Intentionality Theory*

If tracking theories and functional role theories don't work, then what does? Mendelovici (2018) and others have proposed the *phenomenal intentionality theory*, or PIT, and I claim with them that this theory succeeds where tracking and functional role theories fail. The main claim of PIT is that all originally intentional states arise from phenomenal consciousness, and this makes PIT far more successful at accounting for the intentionality of perceptual color representation than tracking theories or long-arm functional role theories: PIT predicts that the content of my representation of sky-blue is the very edenic color which we observe through introspection. In other words, if phenomenal consciousness grounds intentionality, then intentional phenomenal states like perceptual color representation are easily accounted for.

The challenging cases for PIT are the intentional states which are not obviously phenomenal in character, such as thoughts, standing states, and nonconscious occurrent states. Here, I will show only how PIT might handle standing states such as standing beliefs or desires, but Mendelovici (2018) gives a far more detailed analysis of PIT's success in these areas.

Some of the main mental states which seem to be intentional are *standing states*, which are "mental states that need not be used, entertained, or otherwise active at the time at which they are had" (Mendelovici, 2018, 161). Beliefs and desires are common examples of standing states. An example due to Pitt (2016) is that of a dreamless sleeper: suppose I ask someone named Bailey if Trump's legacy will be divisive.[2] Bailey thinks for a moment and says "yes, I do believe Trump's legacy will be divisive." Then, Bailey goes to sleep, and has no dreams. It would be very natural to say, even while Bailey is dreamlessly asleep, that she believes that Trump's legacy will be divisive. In this sense, Bailey's belief seems to persist throughout

[2] I use here the proposition "Trump's legacy will be divisive" which is due to Coleman (2022), whose discussion of both Pitt's view and Crane's view is fantastic.

unconsciousness. Crane (2014) takes this persistence to be one of the defining features of belief: "belief is not just a matter of taking something to be the case for the duration of (e.g.) a perceptual experience. Rather, it is essential to beliefs that they persist and through changes of in current consciousness" (p. 271). But, if this is true, then it seems to be the case that conscious experience is *not* essential to belief, and that beliefs do not require consciousness to occur. Therefore, this would perhaps constitute a challenge for PIT.

Pitt (2016) and Mendelovici (2018) provide a potential response, which amounts to claiming that there are no genuinely intentional standing states. Mendelovici calls this view eliminative dispositionalism, and the main idea is this: when I say of sleeping Bailey that she believes Trump's legacy will be divisive, I am not speaking literally. Instead, to borrow Pitt's example, saying Bailey believes Trump's legacy will be divisive is like saying that Bailey sings well (Pitt, p. 124). When I say of sleeping Bailey that she sings well, I do not literally mean that she currently is singing, and that she is doing a good job at it. I rather mean that *when* Bailey sings, she is good at it. The idea is that this is also the case for belief attributions and attributions of standing intentional states in general. This makes attributions of standing intentional states *dispositional*: Bailey is *disposed* to believe that Trump's legacy will be divisive, just as she is disposed to sing well. Pitt provides another useful analogy for this claim:

> Having a persisting belief in your brain is like having a photograph on the hard drive of your computer. There are not really any photographs in your computer. (Look closely; you will not see any.) What there are are dispositions to (re)produce photographs (on the computer screen, on printer paper, or whatever). (p. 124)

If this is right, then standing intentional states such as persisting beliefs or desires do not actually exist. Instead, there are dispositions to have occurrent intentional states, which are necessarily conscious.

PIT is therefore able to handle the troublesome cases of standing states, and, as Mendelovici (2018) argues, it can do so also with thoughts and nonconscious occurrent states. Given PIT also succeeds with perceptual color representations, and that tracking theories and functional role theories do not, PIT stands as the most plausible theory of intentionality. So, our most plausible theory of intentionality is that it arises from phenomenal consciousness, which, I have argued, physicalism cannot account for. This means that physicalism's failure as a theory of mind is not restricted to consciousness: its failure in accounting for consciousness also constitutes a failure in accounting for intentionality. Since consciousness and intentionality are the two defining features of the mind, this makes physicalism an extremely untenable theory of mind.

3 Science, Knowledge, and Causation

The previous sections argued against physicalist theories of mind, showing that they lack the explanatory capacities required to account for the mind. As mentioned in the introduction, these arguments are aimed directly at philosophers of mind. In this section, though, I want to anticipate some of the worries someone might have about the scientific and epistemic implications of my conclusions. In short, I will answer here the question of "What questions can science answer, and what questions can it never hope to answer?" As we will see, the answer will depend on whether we double-down on physicalist theories or whether we reject them in favor of dualism, and the difference between these two approaches make dualism more favorable than physicalism.

3.1 How Physicalism's Failures Impact Present-Day Science

The hard problem of consciousness is the main problem for physicalism. It tells us that science cannot explain why phenomenal consciousness exists. As I argued above, this explicitly

weighs on physicalism as a theory of mind. How much does it weigh on the empirical sciences, such as psychology or neuroscience? Well, first of all, recall that the hard problem of consciousness is distinct from many *easy* problems of consciousness. The easy problems involve questions like "How do we deliberately control our bodily movements?" or "How do we change our focus of attention from one thing to another?" But these problems are clearly scientifically solvable because they are problems of explaining particular mental functions; in Chalmers's (1995) words, "All of them are straightforwardly vulnerable to explanation in terms of computational or neural mechanisms" (p. 2). So, the questions posed by the easy problems are questions which science can answer. The hard problem, however, is not a problem of explaining any mental function. It is the problem of answering the question "How does consciousness arise?" or "Why does consciousness exist?" The failures of physicalism discussed in the first section of this paper shows that science cannot hope to answer these questions.

Another easy problem that Chalmers (1995) mentions is "the ability of a system to access its own internal states" (p. 2). This is important because some cognitive scientists distinguish between different forms of consciousness. Block (1996), for instance, famously demarcated phenomenal consciousness from access consciousness (p. 3). Block takes phenomenal consciousness to involve the experiential aspects of "what it's like" to be something, in the sense discussed above, and it is the kind of consciousness that Chalmers shows to be scientifically inexplicable. Block takes access consciousness, however, to merely involve our ability to access certain facts or mental contents (p. 3). I won't comment here on whether this distinction is legitimate, but Chalmers clearly indicates that questions of access are easy, in the sense of being scientifically explicable (1995, p. 2). Again, this is because access is a mental function, and mental functions are explicable in terms of computational or neural mechanisms. So, access—

whether it is a part of consciousness or not—is an easy problem. As a result, questions of access are questions which science can answer.

3.2 How Physicalism's Failures Impact Causality

Now, one might reasonably have concerns that the failures of physicalism bear directly on matters of mental-physical causation, that is, how aspects of physical world causes certain events of subjective phenomenal experience (as the breaking of my wrist causes me to experience pain) or how aspects of my experience cause certain physical events (as my experience of pain causes a noise like "ow!" to come out of my mouth). After all, it seems that the empirical sciences' notion of causal interaction is only intelligible when it occurs between two physical substances, such as the Earth causing an apple to fall to the ground. So, if physicalist theories fail, how do scientists make sense of mental-physical causation?

3.2.1 *Doubling-Down on Physicalism*

This question is complicated, but one possible answer appears upon close examination of the argument presented in the first section. Recall that the argument there was that physicalist theories fail, not that the mind is non-physical. This form of argumentation explicitly leaves open the option that the mind *is in fact* physical—that some physical process *does in fact* give rise to consciousness—but that we can never explain why this is so. If this is the case, then mental-physical causation is a non-problem, as it is just a particular kind of physical-physical causation. The problem, of course, is that it seems very odd to suppose that there is a physical process of which humanity is doomed to remain completely ignorant. This is even more clear when we remember that the physical process that gives rise to consciousness (if one exists) occurs in systems where we *do* have very good empirical access, namely mammalian brains. In other words, if we answer the question of mental-physical causation in this way, then we would be

forced to accept the fact that the brain—an organ which thousands of scientists study every day—operates some physical process which science can never hope to uncover. But this is not all. If we take this route, then science's ability to solve the easy problems of consciousness are also weakened. Science may, for instance, be able to give us a mechanical function which lets a system deliberately control its own behavior. But since phenomenal experience plays a role in *our* deliberate control of *our* behavior, if the causal chain from phenomenal consciousness to other physical processes is fundamentally unknowable by us, then the details of how our conscious deliberation results in behavioral control will be unknowable, too. So, answering the question of mental-physical causation in this way would put some upper bound—though a large bound—on scientific knowledge of both phenomenal consciousness itself and the mental processes which interact with it.

3.2.2 *Dualism*

If we are more inclined to say that science can discover everything that there is to know about the brain, at least in principle, then we are forced to answer the question of mental-physical causation differently. If we take this route, then we will be forced to say that phenomenal consciousness is not the result of a physical process; its resistance to physical explanation would, in this case, imply its non-physical nature. The way to account for consciousness would therefore be to take it as fundamental or to posit some non-physical substance process which results in consciousness. This is dualism, which would certainly make many scientists and philosophers dissatisfied. However, if we take consciousness as fundamental, we end up with a kind of dualism which Chalmers calls "innocent" because of its similarity to the theoretical output of *physics*. Our best theories of physics posit certain fundamental entities—quarks, leptons, an electromagnetic field, etc.—which interact with each-other through posited, fundamental

interaction principles such as "a photon affects the electromagnetic field in this way, which affects other photons in this way." The question for those inclined towards physicalism, then, is whether there is any good reason to suppose that the positing of fundamental, non-physical entities with fundamental interaction principles is any different from the kinds of positing which physicists do. If the existence of fundamental entities and interaction principles is the best explanation for the phenomena we observe—and, as our discussions in the previous sections reveal, it is—it is hard to see why we should rule this option out.

Chalmers (1995) points out that this approach suffers from one of the same problems as the other, namely that it does not really explain why consciousness exists because it merely posits it as fundamental (p. 15). But, as he notes, "Nothing in physics tells us why there is matter in the first place, but we do not count this against theories of matter" (p. 15). What this shows is that even naturalistic metaphysical theories are not burdened with explaining the existence of things which it takes as fundamental. So, if extending our ontology is necessary to account for consciousness, then this move does not directly negatively impact scientific theories. This means that the kind of explanatory failure of dualism is different from the kind of explanatory failure for physicalism. While the kind of failure that occurs in dualist theories of mind also occurs in physics' theories of matter, the kind of failure that occurs in physicalist theories is not encountered anywhere else. We also saw how, if we take consciousness to be physical but inexplicable, then our ability to solve the easy problems of consciousness are also hampered. This was because the process that results in conscious experience plays a role in the processes described in the easy problems. So, if we cannot know what process gives rise to consciousness, then we cannot understand every detail of the causal chain of other mental processes. However, if we take consciousness to be fundamental, then we do not run into this issue—where doubling-

down on physicalist theories negatively impacts science's ability to answer the easy problems of consciousness, accepting the innocent version of dualism presented here does not.

4 Conclusion

In section 1, we discussed how the hard problem of consciousness presents a problem for physicalism just as the mind-body problem presents a problem for dualism. The similarity of these two problems is enough to conclude that physicalism is just as successful as dualism is; since dualism is (generally speaking) taken to fail as a result of the mind-body problem, physicalism also ought to be taken to fail as a result of the hard problem of consciousness.

The distinguishing features of the mind are consciousness and intentionality. Due to Chalmers's (1995) hard problem of consciousness, physicalism cannot account for phenomenal consciousness. Then, since phenomenal consciousness provides the most plausible account of original intentionality, physicalism cannot account for intentionality either. This means physicalist theories of mind fail almost completely and have no advantages over dualist theories of mind. Additionally, if we double-down of physicalist theories, then science's ability to solve the easy problems of consciousness are hampered. This makes dualist theories more desirable.

Very nice work! I can't say I'm persuaded to abandon physicalism just yet, but you're giving the view an extra run for its money by making even intentionality seem like a problem for physicalists (we like to think we've had success on that front!). I'll have to think more about what to make of the idea that intentionality is grounded in consciousness. Overall, this was a pleasure to read and has given me plenty of food for thought.

160/160

References

Block, N. (1996). How not to find the neural correlate of consciousness. In Branquinho, J., *The foundations of cognitive science* (pp. 1-10). Oxford University Press.

BonJour, L. (1998). *In defense of pure reason: A rationalist account of a priori justification*. Cambridge University Press.

Chalmers, D. (1995). Facing up to the hard problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200-219. Used electronic version: https://consc.net/papers/facing.pdf.

Chalmers, D. (1996). Can consciousness be reductively explained? In Chalmers, D., *The conscious mind: In search of a fundamental theory* (pp. 93-105). Oxford University Press USA - OSO. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/osu/detail.action?docID=272854.

Coleman, S. (2022). The ins and outs of conscious belief. *Philosophical Studies* 179(2), 517-548.

Crane, T. (2014). Unconscious Belief and Conscious Thought: 2012. In *Aspects of Psychologism* (pp. 261-280). Harvard University Press.

Mendelovici, A. (2018). *The phenomenal basis of intentionality.* Oxford University Press.

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4), 435–450. https://doi.org/10.2307/2183914.

Pitt, D. (2016). Conscious belief. *Rivista Internazionale di Filosofia e Psicologia* 7(1). 121-126.

Searle, J. R. (1983). *Intentionality, An essay in the philosophy of mind*. Cambridge University Press.