Précis of *Consciousness and the Prospects of Physicalism*

Oxford University Press, 2011

Derk Pereboom, Cornell University

Philosophy and Phenomenological Research, 86 (2013), pp. 715-27 and 753-64.

Penultimate Draft

Consciousness and the Prospects of Physicalism has three parts. The first (Chapters 1-4) develops a response to the knowledge and conceivability arguments against physicalism, one that features the open possibility that introspective representations represent mental properties as having features they actually lack. The second part (Chapters 5 and 6) proposes a physicalist version of a Russellian Monist answer to these arguments, the core of which is that currently unknown intrinsic physical properties provide categorical bases for known physical properties and also yield an account of consciousness. The third part (Chapters 7 and 8) defends a nonreductive account of physicalism about the mental, according to which the relation between the mental and the microphysical is constitution, where this relation is not explicated by the notion of identity.

The challenge to the anti-physicalist arguments set out in the first part of the book begins with the observation that phenomenal states have characteristic phenomenal properties, and that it at least at first intuitive for many of us that:

> (i) Introspective modes of presentation represent a phenomenal property as having a specific qualitative nature, and the qualitative nature that an introspective mode of presentation represents a phenomenal property as having is not among the features that any physical mode of presentation would represent it as having.

It is also initially intuitive that:

> (ii) The introspective mode of presentation *accurately represents* the qualitative nature of

the phenomenal property. That is, the introspective mode of presentation represents the phenomenal property as having a specific qualitative nature, and the attribution of this nature to the phenomenal property is correct.

A Lockean way to characterize the qualitative natures that introspective modes of presentation represent phenomenal properties as having is by way of resemblance to modes of presentation. Mary's introspective representation of her phenomenal-red sensation presents that sensation in a what-it-is-like-to-sense-red way, and it is intuitive that a qualitative nature that resembles this what-it-is-like mode of presentation is accurately attributed to the sensation's phenomenal property. Or else, in deference to concerns about the cogency of such resemblance characterizations, it might be specified instead that this qualitative nature is as the introspective mode of presentation represents it to be.

Given these claims about what is at least initially ntuitive, an advocate of the knowledge argument can account for the force of the knowledge argument in the following way. When Mary leaves the room and sees the tomato, she comes to have the belief:

(A) Seeing red has R,

where the concept 'R' in this belief is the phenomenal concept that directly refers to phenomenal redness, i.e., to phenomenal property R. The qualitative nature of phenomenal redness is accurately represented introspectively by way of the *what-it-is-like-to-sense-red* introspective mode of presentation. The argument is profitably read as assuming that every truth about the qualitative nature that an introspective mode of presentation accurately represents a phenomenal property as having would need to be derivable from a proposition detailing only features that physical modes of presentation represent the world as having (an assumption I don't challenge). However, (A) is not derivable from such a proposition. Thus not every truth about the qualitative

nature that an introspective mode of presentation accurately represents a phenomenal property as having is so derivable. Therefore physicalism about phenomenal properties is false.

One might challenge this version of the knowledge argument by taking issue with one or both of the claims about what is intuitive just listed. The one I develop a case against in the first four chapters is (ii), the claim about the accuracy of introspective representation. The idea is that, given the supposition of (i), it is an open possibility that introspective representation is inaccurate in the respect that it represents phenomenal properties as having qualitative natures they do not in fact have; that is, it is an open possibility that the *qualitative inaccuracy hypothesis* is true. For example, upon seeing the red tomato, Mary introspectively represents the qualitative nature of phenomenal redness in the *what-it-is-like-to-sense-red* way, and it is an open possibility she then represents phenomenal redness as having a qualitative nature that it actually lacks.

The seriousness of this open possibility can be supported by an analogy with our perceptual representations of secondary qualities. Our visual system represents colors as having certain qualitative natures, and it is an open possibility, widely regarded as actual, that colors actually lack them. It may be that relevantly similar mechanisms to those involved in visual color perception feature in introspective phenomenal property representation, and this would explain how qualitative inaccuracy hypothesis about our representations of phenomenal properties could be true. It might well be that part of what explains qualitative inaccuracy about color representation is that it is causal, and this allows for a general discrepancy between the real nature of color and how we represent its qualitative nature. It's open that introspective representation of phenomenal color is similarly causal, and that this gives rise to an analogous general discrepancy.

On the supposition that the open possibility is in fact realized, how should we describe what happens when Mary leaves the room and sees the red tomato? She now has a belief of the form:

(A) Seeing red has R.

Consider first the initially plausible proposal that the concept 'R' in this belief refers to a property with the qualitative nature accurately represented by the introspective what-it-is-like-to-sense-red mode of presentation. On our open possibility, phenomenal redness has no such qualitative nature, and the belief will be false. Then Mary's coming to believe (A) does not amount to her acquiring a new true belief about the qualitative nature of phenomenal redness. belief. Next, consider the contrasting proposal that phenomenal property R is introspectively misrepresented and lacks such a qualitative nature, and is wholly physical. We can suppose that this belief is then true, but while she was in the room Mary already had this belief, or was able to derive it from the true beliefs she already had, and thus she again does not acquire a new true belief. Thus either the belief (A) about phenomenal redness is not true, or it turns out not to be new.

Now consider the open possibility just as a hypothesis about how things might turn out to be. Does this give us a reason to believe that Mary hasn't acquired a new true belief? In my estimation, the open possibility is serious enough to provide us with such a reason. In fact, my sense is that this possibility is sufficiently serious to preclude rational conviction that Mary does acquire a new true belief, and herein lies the challenge to the knowledge argument.

One might suspect that the problem for a physicalist explanation has merely been shifted from accounting for phenomenal states and their properties to accounting for their introspective phenomenal modes of presentation. On the qualitative inaccuracy hypothesis, Mary's

phenomenal state lacks a qualitative nature that is accurately represented by her introspective phenomenal mode of presentation – call it MPR. So Mary does not learn that the sensation has a property of this particular sort. But it would seem that she does learn something about how MPR presents this sensation. Since MPR presents the sensation phenomenally, she would appear to learn something about a phenomenal property of MPR – in particular, something about its essential property of presenting red sensations in the what-it-is-like-to-sense-red phenomenal way. It seems initially plausible that Mary cannot derive the corresponding phenomenal truth about MPR from her microphysical base.

In response, there is no less reason to think that the qualitative inaccuracy hypothesis applies to introspective representations of phenomenal modes of presentation than to introspective representations of phenomenal states. It's thus also an open possibility that Mary introspectively represents MPR's essential phenomenal property as having a qualitative nature it really lacks. But then, despite how MPR is introspectively represented, it might be that while she is still in the room Mary can derive every truth about its real nature from her the microphysical base. So even though while she is in the room, Mary has never introspectively represented MPR, it is an open possibility that she can derive every truth about it. The same type of point can be made for any further iteration of introspective representations of introspective phenomenal modes of presentation. Any information about the real nature of any phenomenal entity would then be derivable from the information Mary has before leaving the room.

The qualitative inaccuracy strategy also yields a challenge to the conceivability argument. I focus on David Chalmers's version, which involves the following notions: 'P' is a statement that details the complete microphysical truth about the actual world; 'T' is a "that's all" provision, so that 'PT' specifies all the physical truths about the actual world with the provision

that there are no further truths (that is, other than those entailed by those physical truths); and 'Q' is an arbitrary phenomenal truth. Statement S is ideally conceivable when it is conceivable on ideal rational reflection; S is positively conceivable, for instance, when we can form a mental picture of a scenario in which S is true. S is primarily possible just in case it is true in some world considered as actual, and S is secondarily possible just in case S is true in some world considered as counterfactual. S is primarily conceivable just in case S can be conceived as true in some world considered as actual, or alternatively, since considering-as-actual is an a priori matter, S is primarily conceivable just in case the subject can't rule out S a priori.[1] According to Russellian monism, underlying the (familiar) physical properties, which are all dispositional and relational, there are categorical and intrinsic phenomenal properties, or else categorical and intrinsic protophenomenal properties, so-called because while not phenomenal properties themselves, they nonetheless account for them (more on this below). Here then is the argument:

(1) 'PT and ~ Q' is ideally, positively, primarily conceivable.

(2) If 'PT and ~ Q' is ideally, positively, primarily conceivable, then 'PT and ~ Q' is primarily possible.

(3) If 'PT and ~ Q' is primarily possible, then 'PT and ~ Q' is secondarily possible or Russellian monism is true.

(4) If 'PT and ~ Q' is secondarily possible, materialism is false.

[1] Thus 'PT and ~ Q' is primarily conceivable for Mary if she can't rule it out a priori. For understanding Chalmers's argument, it's important to see that the can rule it out a priori if she can a priori derive 'Q' from 'PT.'

(5) Materialism is false or Russellian monism is true.[2]

My challenge is to Premise (1). As Chalmers sees it, rejecting this premise involves claiming that an ideal reasoner could derive a priori the arbitrarily selected actual phenomenal truth 'Q' from 'PT,' supposing she has the minimal information required to ensure adequate possession of the phenomenal concepts involved in representing 'Q.' He argues that it is strongly intuitive that this claim is false. By contrast, in his view a truth like 'water exists' can be derived a priori from 'PT', and thus 'PT and there is no water' will not be ideally, positively, primarily conceivable. But the premise can be challenged as follows. Partly because the qualitative inaccuracy hypothesis is an open possibility, the analysis of phenomenal concepts discloses a conjunction of conditionals of the following sort:

(P1+) If a world is actual in which experiences are qualitatively exactly as we introspectively represent them to be, then:

the concept 'phenomenal red' correctly applies to phenomenal redness as it is introspectively represented, and

(P2+) If a world is actual in which no experiences instantiate phenomenal redness as it is introspectively represented, but there is a unitary property that is the normal cause of their introspectively appearing phenomenally red, then:

the concept 'phenomenal red' correctly applies to the property that is the normal cause of the introspective appearance of phenomenal redness (where 'the normal

[2] David Chalmers, "Does Conceivability Entail Possibility?" in *Conceivability and Possibility*, ed. Tamar Gendler and John Hawthorne, Oxford: Oxford University Press, 2002, pp. 145–200.

cause of introspective representations of phenomenal redness' functions merely as a reference-fixer), and

(P3+) If a world is actual in which no experiences instantiate phenomenal redness as it is introspectively represented, but there are many different sorts of causes of their introspectively appearing phenomenally red, and there are no salient similarities among the intrinsic properties of these causes, then:

the concept 'phenomenal red' correctly applies to whatever properties cause (or could cause) instances of the introspective appearance of phenomenal redness.

Thus in a scenario microphysically just like ours (and 'T' holds) but without instantiated phenomenal properties whose qualitative natures are accurately represented by introspective modes of presentation, phenomenal properties might well nevertheless be instantiated. And if such a scenario was actually realized, there would be no less reason to believe that 'PT and ~ Q' would be ruled out by ideal a priori reasoning than to believe that 'PT and no physical objects are red' or that 'PT and there is no water' would be so ruled out. Consequently, the status of the ideal, positive, primary conceivability of 'PT and ~ Q' will not differ from that of 'PT and there is no water' and 'PT and no physical objects are red.' Thus, since the qualitative inaccuracy hypothesis is an open possibility, Premise (1) is insecure.

The subject of chapters 5 and 6 is the Russellian Monist response, whose core idea is that currently (but not inevitably) unknown or at least incompletely understood intrinsic properties provide the categorical bases for the known physical dispositional properties, and would also yield an account of consciousness. While there are nonphysicalist versions of this view, some are amenable to physicalism. In Chalmers's terminology, the variants that are potentially

physicalism-friendly propose that the fundamentally intrinsic properties are protophenomenal, that is, properties that are not phenomenal but nonetheless explain the instantiations of phenomenal properties. The resulting type of physicalism has an advantage over the kind discussed in the first four chapters, for the reason that it can preserve the intuitive claim that phenomenal properties really possess the qualitative natures we introspectively represent them as having.

For many, it seems at least initially reasonable to conjecture that consciousness is not a fundamental phenomenon, and that there are more fundamental features of reality that underlie and explain it. Current physics encourages the hypothesis that the fundamental features of reality are physical; candidates include particles, forces, and quantum fields. But there are serious considerations, such as the knowledge and conceivability arguments, that tell against the view that anything physical of the sort we now understand can account for consciousness. This situation gives rise to the thought that the account must consist at least in part in presently unknown fundamental features of reality. Add to this that the history of philosophy has shown a strong predilection for ontological monism, that is, for maintaining that the world has fundamental features only of a single kind – materialism and idealism are cases in point. These motivations give rise to a position on which not only consciousness, but also the sorts of physical features encountered in current physics, is grounded in underlying fundamental features of a single kind. This is Russellian Monism, named for one of its proponents, Bertrand Russell.[3]

One more specific Russellian Monist proposal involves the notions of dispositional and categorical properties. Dispositional properties are essentially tendencies to produce certain

[3] Bertrand Russell, *The Analysis of Matter* (London: Kegan Paul, 1927); the classic passage is on p. 384.

effects, and while categorical properties may have tendencies to produce effects, their natures do not consist only in such tendencies. Many find it intuitive that categorical properties are needed to account for dispositional properties; for example, a ball's disposition to roll requires an explanation, and it is explained by its categorical properties of spherical shape and rigidity. The more specific Russellian monist proposal then is this: the most basic properties current physics reveals are all dispositional, while it leaves us ignorant of the categorical properties needed to explain them, and these unknown categorical properties account for consciousness. An electron's negative charge, for instance, is one of those basic physical properties, and it is a disposition to repel other particles with negative charge and to attract particles with positive charge. This dispositional property must have a categorical basis, and it, the Russellian Monist hypothesizes, is the kind of property that can also account for our consciousness. Russellian Monists have proposed a range of such more fundamental but yet undiscovered properties – from conscious properties of microphysical entities, to properties of such entities similar enough to paradigmatic physical properties to qualify as physical themselves but yet different enough to explain consciousness, to properties unlike any we've encountered, but capable of explaining consciousness.

The particular version of Russellian Monism I set out is one according to which the yet-to-be discovered properties crucial to explaining consciousness are of the second sort, close enough in kind to our paradigmatic physical properties to count as physical. What distinguishes my formulation from others is that in mine these currently unknown properties are not only categorical but also intrinsic – that, is, non-relational – in a certain demanding sense. The idea of this sort of intrinsic property derives from Leibniz's discussion of the Cartesian view according

to which matter consists just in extension in three spatial dimensions.[4] Leibniz contends that the extension of the sphere can be analyzed as, or reduces to, the plurality, continuity, and coexistence of parts of the sphere, each of which is a purely extrinsic property of these parts, that is, an extrinsic property with no intrinsic aspects; (in this context, I treat 'extrinsic' and 'relational' as synonymous).[5] *Being one of a collection of more than one thing*, *being spatially continuous with other things*, and *coexisting temporally with other things* are purely extrinsic properties of their bearers. So it may be that P is an intrinsic property of X, while P is not in a sense fundamentally intrinsic to X, or, as James van Cleve points out, in Kant's terminology, *absolutely* intrinsic to X.[6] This is the case when X's having P can be analyzed as, or reduces to,

[4] Leibniz to deVolder, April 1702, in G. W. Leibniz, *Philosophical Papers and Letters*, ed. L. E. Loemker (Dordrecht, The Netherlands: D. Reidel, 1969), pp. 526–27.

[5] *Being wise* is an example of an extrinsic property that isn't purely extrinsic. Sophie's being wise is an extrinsic property of hers because it involves a relation to a comparison class, but it also includes an intrinsic aspect – having a certain type and level of intelligence. *Being wise* is an example of an extrinsic property that isn't purely extrinsic. Sophie's being wise is an extrinsic property of hers because it involves a relation to a comparison class, but it also includes an intrinsic aspect – having a certain type and level of intelligence. *Being wise* is thus a complex property that has at least one extrinsic and one intrinsic aspect, and as a result it isn't purely extrinsic. But this raises the possibility of extrinsic properties with no intrinsic aspects – purely extrinsic properties.

[6] Immanuel Kant, *Critique of Pure Reason*, A277/B333, tr. Paul Guyer and Allen Wood (Cambridge: Cambridge University Press, 1987). James van Cleve, "Inner States and Outer Relations: Kant and the Case for Monadism," in *Doing Philosophy Historically*, ed. Peter H.

X's parts having properties Q, R, S . . ., and these properties are purely extrinsic properties of these parts. Correlatively, when P *can* be analyzed as or reduces to purely extrinsic properties of these parts, it is instead, in Kant's terminology, merely comparatively or relatively intrinsic.

But it's best to avoid the notions of analysis and reduction in characterizing these properties. Even if for general reasons supporting anti-reductionism, properties of the whole fail to be analyzable in terms of or to reduce to properties of the parts, an intrinsic property of the whole could still be merely comparatively intrinsic.[7] We can instead appeal to the notion of constitution without identity (my version of material constitution is spelled out in Chapter 7 -- see below; this account can be modified to accommodate property constitution):

> P is an *absolutely intrinsic* property of X just in case P is an intrinsic property of X, and P is not even partly constituted by purely extrinsic properties of parts of X.

By contrast,

> P is a *comparatively intrinsic* property of X just in case P is an intrinsic property of X, and P is wholly constituted (at some level) by purely extrinsic properties of parts of X.

Leibniz then argues, in effect, that every substantial entity has at least one absolutely intrinsic property, and so, contrary to Descartes's proposal, extension alone is implausibly constitutive of material substance. One component of Russellian Monism can explained along these same lines:

Hare (Buffalo, NY: Prometheus, 1988), pp. 231–47.

[7] These definitions are revisions of those in the book. Thanks to Chase Wrenn, Ralf Bader, and Nico Silins for comments that occasioned the revisions. Silins suggested that I use the my notion of constitution instead of reduction or necessitation.

the properties that physics reveals to us are all extrinsic or only comparatively intrinsic, and so there must be currently unknown absolutely intrinsic properties that underlie them.

Chalmers's Russellian Monist thought is that one can ideally, positively, primarily conceive 'PT and ~ Q' (that is, conceive it as true in some world considered as actual) only because one is conceiving just extrinsic/dispositional properties on the physical side. We can now suggest that if 'P' were replaced with an embellished 'P*' that includes concepts that allow for representation of the natures of the currently unknown absolutely intrinsic properties, the resulting 'P*T and ~ Q' would not be ideally, positively, primarily conceivable. For although 'Q' – let's suppose, 'Mary senses red at time t' – is not a priori derivable from 'PT,' this claim about Mary's phenomenal experience would be a priori derivable from 'P*T.' The Russellian Monism that ensues has versions on which the natures of the absolutely intrinsic properties are phenomenal, as in Galen Strawson's micropsychism, or else not phenomenal but protophenomenal, as Chalmers advocates. On the version of protophenomenalism I explore, these properties are similar enough to paradigmatic physical properties to count as physical themselves. It would be theoretically advantageous if the hypothesized absolutely intrinsic properties provided explanations for both phenomenal properties and the properties specified by current microphysics. I argue that it's open that a protophenomenalist version of Russellian Monism is equipped for this twofold task.

Might we ever come to possess concepts that allow us to represent the natures of protophenomenal properties that satisfy these conditions? Chalmers is cautiously optimistic. In a slightly different context, Colin McGinn is skeptical. By contrast with acquiring concepts that facilitated past major theoretical shifts in science, such as the advance to relativity theory, this we cannot achieve; "what we need is a perspective shift, not just a paradigm shift—a shift not

merely of world view, but of ways of apprehending the world. We need to become another type of cognitive being altogether."[8] But Chalmers, like Thomas Nagel, thinks it is open that our cognitive and imaginative capacities are up to forming the required sort of concept, and I side with them.[9]

The third theme of this book is nonreductive physicalism about the mental, the topic of chapters 7 and 8. In the version I defend, the core reason for nonreductivism is not methodological or pragmatic but metaphysical. Natural kinds in psychology are not identical to natural kinds in physics because psychological causal powers are not identical to microphysical causal powers. The fact that psychological kinds are multiply realizable at the level of microphysical kinds yields the important clue as to why this is so. The nonreductivism I set out departs from other nonreductivisms in that it rejects the token identity of psychological and microphysical entities of any sort—including causal powers. The deepest relation between the psychological and the microphysical is constitution, where this relation is not to be explicated by the notion of identity.

In my conception, constitution is a grounding relation between concrete physical entities; they might be states, events, property instances, or causal powers. Suppose x and y are concrete physical entities. The *made up of* relation is asymmetric, irreflexive, and directed, so that the less fundamental is made up of the more fundamental, while its core is primitive. The *made up of* relation is asymmetric and irreflexive: the lattice is not made up of the diamond, and the

---

[8] Colin McGinn, "What Constitutes the Mind-Body Problem," in his *Consciousness and Its Objects* (Oxford: Oxford University Press, 2004), p. 24.

[9] David Chalmers, "Does Conceivability Entail Possibility?"; cf., Thomas Nagel, *The View from Nowhere*, New York: Oxford University Press, 1986.

diamond is not made up of itself. It has a specific direction: the less fundamental made up of the more fundamental. Entities x and y are materially coincident just in case they, at some level, are made out of the same parts. Then,

(C1) x materially constitutes y at t if and only if

(a) y is made up of and materially coincident with x at t;

(b) necessarily, if x exists at t, then y exists at t and is made up of and materially coincident with x at t; and

(c)) possibly, y exists at t and it is not the case that y is made up of and materially coincident with x at t.

The last clause (c) precludes the identity of x and y (on the assumption of the necessity of identity), as does clause (a), since the *made up of* relation is irreflexive.

Lynne Baker's discussion of constitution features a number of counterexamples that would pose a threat to clause (b) of this characterization, the necessitation of the constituted entity by its constitutor.[10] The existence of the dollar bill in my pocket is not necessitated by its cellulose molecule and ink constitutor, for its existence also requires the US Federal Reserve Bank and the laws governing it (or some similar arrangement). For a mental example, on an externalist view about psychological content of the kind developed by Tyler Burge, the existence of a token belief with some specific content will not be necessitated by the existence of its neural or microphysical constitutor, for in an alternative physical and social environment, this same neural or microphysical constitutor would not yield a belief with that content. Phenomena of these kinds can be accommodated by a characterization very close to (C1), in which, on the

[10] Lynne Baker, *The Metaphysics of Everyday Life* (Cambridge: Cambridge University Press, 2007), pp. 11–13, 106–10.

recommendation of Baker's account, (b) is revised to specify that the existence of y is necessitated by the existence of x in an appropriate relational context, and (c) is similarly altered. Suppose 'D' designates the y-favorable circumstances—the relational context required for something to constitute y. Then:

(C2) x materially constitutes y at t if and only if

(a) y is made up of and materially coincident with x at t;

(b) necessarily, if x exists and is in D at t, then y exists at t and is made up of and materially coincident with x at t; and

(c)) possibly, y exists at t and it is not the case that y is made up of and materially coincident with x in D at t.

I argue that a view on which mental states are constituted of lower-level states on this conception can answer the exclusion problem that Jaegwon Kim poses for the nonreductivist.

In the last chapter, I set out a model of the mental that is not functional in the standard sense, that is, one in which the essences of types of mental properties do not consist in their causal relations to sensory inputs, behavioral outputs, and other mental states. Instead, mental properties are identical to broadly physical compositional properties, properties things have solely by virtue of intrinsic features of their parts, either proper or improper, and relations these parts have to one another. This model would secure the causal efficacy of the mental, given a nonreductive view, in a way that the standard sort of functionalism arguably cannot. It would also preserve nonreductivism, since multiple realizability arguments indicate that mental compositional properties would not be essentially neural or microphysical. Given the identities

that it affirms, in a significant respect the position espoused amounts to a compromise with the type-type reductionist views of U. T. Place and J. J. C. Smart.[11]

The realist model inherited from other sciences is plausibly interpreted as explaining the forward-looking causal relata of kinds not simply by way of functional relations, but rather in central cases by properties intrinsic to those kinds – that is, by properties intrinsic to every possible instance of the kind – and in particular by intrinsic properties at the same level as the kinds themselves. In chemistry, the forward-looking causal relata of compounds are explained in part by their compositional properties, chemical properties intrinsic to those kinds of compounds, which each molecule of the compound has by virtue of the intrinsic properties of its component parts and the relations these parts have to one another. In biology, polio symptoms are explained partly by an intrinsic biological property of that kind of disease, being a particular viral infection. Similarly, in this model properties intrinsic to types of mental states that explain their forward-looking causal relata.

This model might be elaborated with the analogy of artifacts of internally complex types, for instance, a ball piston engine, a recent version of the rotary internal combustion engine. Characteristic of this engine is having parts with certain shapes and rigidities, arranged in a particular way. These features comprise a compositional property intrinsic to such an engine, and they are not external functional relations that such an engine stands in. This compositional property is multiply realizable. The parts of the engine can be made of materials of different sorts, as long as these materials can play the right role, for example yielding the requisite shapes

[11] U. T. Place, "Is Consciousness a Brain Process?" *British Journal of Psychology* 47 (1956), pp. 44–50; J. J. C. Smart, "Sensations and Brain Processes," *Philosophical Review* 68 (1959), pp. 141–56.

and rigidities. Similarly, it might be that the heterogeneous physical realizations of a dog's and a human's belief *that this fire is dangerous* exhibit a compositional property of a single type that is intrinsic to this kind of mental state, a compositional property instances of which are the causal powers of this belief. This property would be more abstract than any specific sort of neural compositional property in the sense that it can be realized in distinct sorts of neural systems. It may be that this same compositional property can also be realized in a silicon-based electronic system, and such a system could then have the belief about danger. Imagine that researchers built a silicon-based system that replicated the capacities of and interconnections among neurons in a human brain as much as is physically possible, and then excited it to mimic as closely as possible what happens when a human has this belief about danger. It's an empirical possibility that the resulting silicon-based state would realize that same belief and have an internal structure that is similar enough to the internal structure of the human neural system for both to instantiate the same compositional property.

Thus on the view proposed in these last two chapters, the mental would thus be twice grounded in the physical: by way of constitution in the microphysical and intervening levels and by way of identity with compositional properties that are sufficiently abstract to preclude classification at any level more basic than the mental.

Replies to Daniel Stoljar, Robert Adams, and Lynne Baker

Derk Pereboom, Cornell University

Penultimate Draft

I wish to thank my three commentators, Daniel Stoljar, Robert Adams, and Lynne Baker, for their superb and thoughtful comments.

Stoljar and Adams each raise doubts about the credibility of the qualitative inaccuracy hypothesis. Again, on this hypothesis, we represent phenomenal properties as having certain qualitative natures that they in fact lack. In the proposed open possibility, the specified kind of inaccuracy is universal or general – it is a feature of all human introspective representation of phenomenal properties, and it is in an important respect extensive – phenomenal properties altogether lack certain qualitative natures they are represented as having. Stoljar and Adams each voice the concern that the generality of the proposed inaccuracy, which is required for defusing the knowledge and conceivability arguments, renders the possibility at issue highly unlikely to be actual. In my account, I support the claim that the possibility is instead seriously open by analogy to the Locke-inspired account of our visual representations of secondary qualities, in which the inaccuracy of these representations is general. In the color case, the causal nature of visual representation is crucial to explaining how this can be. By the standard causal theory of visual representations, they are typically caused by external objects (more specifically by their property-instances) and are distinct from them, and they mediate the subject's representing of the object, and this is what allows for the general disparity between the features objects appear to have and those they really have. I contend that this analogy yields reason to believe that the possibility of a general sort of inaccuracy in the case of phenomenal property representation

could really be actual, since this sort of representation might have a similar causal account. Causal theories of representation are well-understood and highly plausible for a wide range of mental representation, and the hypothesis that introspective phenomenal representation is also causal is live. If introspective representation is causal, the causal mechanisms could really be similar enough to those of visual color representation to generate universal disparity in the phenomenal case as well. Stoljar is right to point out that we know what physical objects need to be like to cause color sensations in us, and this helps explain universal disparity there, while in the case of phenomenal property representation we have no analogous knowledge. But I'm only claiming that it's open that there is universal disparity in the phenomenal case, not that it's established, and mere openness doesn't require the knowledge of introspective processes Stoljar specifies.

Stoljar and Adams argue that the fraternity example, in which cold seems to be misrepresented as pain, and the dentist case, in which pain seems to be misrepresented as cold, don't support the qualitative inaccuracy hypothesis insofar as it is meant to feature the general sort of disparity. I don't claim that these examples function as the main support for the hypothesis, and in my view it's rather the plausibility of a causal account of phenomenal property representation and the parallel to visual color representation that are meant to provide its strongest validation. The fraternity and dentist examples have a related, but yet distinct role. It's a given that most of us don't ordinarily believe that we generally represent phenomenal properties as having certain qualitative natures that they in fact lack. How might this be explained on the assumption that the qualitative inaccuracy hypothesis is true? I propose that it would count in favor of this sort of qualitative inaccuracy if our becoming aware of a discrepancy between the real qualitative nature of a phenomenal property and how it is

introspectively represented sometimes but nonetheless very seldom occurs. Then we would have some reason for thinking that such introspective misrepresentation is possible, and we would have a partial explanation for our resistance to this possibility. And awareness of this sort of discrepancy in fact only seldom occurs -- only in cases like the fraternity and dentist examples. As I acknowledge in the book, there are interpretations of these cases that don't involve introspective inaccuracy that involves an inaccurate perception of a qualitative nature, and Adams defends one on which the mistake lies in conceptualization, not in perception. I agree that this interpretation is open as well.

Another feature of introspective representation that might help explain our resistance to the qualitative inaccuracy hypothesis is that we lack ways of checking whether introspective representations are inaccurate in this way. For external sensory representation, we often do have readily available ways of checking what is represented that are independent of the representation under scrutiny. One might, for instance, measure the Müller-Lyer lines to determine whether one's visual representation of them as having different lengths was correct. But in the phenomenal case, performing an analogous sort of check is not readily available. This limitation yields an explanation for why we would at most be only infrequently aware of discrepancies between the real qualitative natures of phenomenal properties and how they are introspectively represented, which helps account for our resistance to the possibility of phenomenal qualitative inaccuracy consistent with its actually existing. Furthermore, given the scarcity of means of checking the accuracy of such representations, there would be little noticeable difference between having an introspective experience in which we represented phenomenal properties as possessing qualitative features they actually lack, and having one in which phenomenal properties were represented accurately. So what we do and do not notice in having introspective

experience, all by itself, will not adjudicate whether we misrepresent the qualitative natures of phenomenal properties.

In further support for the plausibility of the qualitative inaccuracy hypothesis, I point out that its truth is consistent with certain claims about the correctness of introspective representation. To use Stoljar's example, even if introspective representation inaccurately represents the qualitative nature of being in an itchy phenomenal state, still it may be that a belief *that I am in* an itchy phenomenal state, a belief that is formed directly on the basis of my qualitatively inaccurate introspective representation, is generated by a mechanism that is very reliable. There might only very infrequently be a discrepancy between which phenomenal states I introspectively represent myself as being in and those I am actually in. For us introspective representation might sort phenomenal state and property instances very accurately, and do so directly on the basis of introspective representation of them -- while at the same time phenomenal properties lack the qualitative natures we introspectively represent them as having. This is in accord with the Lockean visual color representation analogy. Although in my visual experience I systematically misrepresent the qualitative nature of color, directly on the basis of such visual experience I am able tell, reliably, which color things are. In this respect the misrepresentation that the qualitative inaccuracy hypothesis proposes is limited.

At this point we might ask: how plausible must the qualitative inaccuracy hypothesis be to count as successful? This question is relevant to the concerns that both Stoljar and Adams raise. As I see it, the original dialectical context is this: the knowledge argument and the conceivability argument require that the qualitative inaccuracy hypothesis be ruled out. If it really might be that introspective phenomenal representation is qualitatively inaccurate, and if it hasn't been shown that this really might not be so, the soundness of these arguments is in

question. But at this point one might abandon this original dialectical context, and ask whether physicalism, given the total evidence, is more plausible than its falsity. In this new type of context, we can ask: how credible must a physicalistic account of phenomenal properties be for physicalism to be more plausible than not? One might think, for instance, that such a physicalistic account of phenomenal properties itself would then need to be more plausible than not.

My sense of the task for such an account, given this epistemic objective for physicalism, does not require that it, considered in isolation from more general reasons to accept physicalism, be more plausible than not. More specifically, when we restrict the evidence to our introspective experience of phenomenal properties and leave out evidence from the history of success of physical explanations in science, the credibility of the physicalistic account need only be modestly high, and in particular it need not be more plausible than not. It would only be in conjunction with such broader evidence that physicalism would be more credible than its falsity. In the book I don't take on the task of assessing the inductive argument for physicalism from the success of physical explanations, which I think is a difficult issue, and so I don't venture an assessment of the overall credibility of physicalism.

An example from Plantinga's discussion of the problem of evil illustrates this epistemic situation. Suppose I know that Feike is a Frisian, and the only evidence I have pertinent to his being able to swim is my awareness of the statistical fact that only one in ten Frisians can swim. Then my rational credence that he can swim is arguably just .1. But suppose now that in addition I swim with Feike every morning. Now it's a (near) certainty for me that he can swim. In this example, we can say that my experience of Feike swimming bears the main burden of raising my

credence that Feike can swim to a high level. This burden cannot be carried by the statistical evidence I have.[12] So if consideration of introspective evidence from phenomenal states alone licenses only modest rational credence that physicalism is true about the phenomenal, the broader evidence could still raise the credence of physicalism about the phenomenal, and more generally, to a much higher level. My sense of the current prospects for physicalism is that we can expect a restricted account to be only be modestly credible, and that the bulk of the burden will need to fall on broader evidence from the history of science. This picture, at least as one of how things currently stand, coheres with the sensibilities of most physicalists. Their confidence that physicalism is true is due largely to considerations that derive from the history of success of physical explanations in science, and not to a high degree of confidence in a physicalist account of phenomenal consciousness considered independently of these broader considerations.

Stoljar objects that I countenance only eliminativism and primitivism about phenomenal properties, and the defender of the knowledge and the conceivability arguments can claim that there are other options. In particular, the defender of these arguments can reject primitivism and still maintain her commitments. I agree with this last claim. I don't think that the knowledge and conceivability arguments depend on primitivism, but just on the weaker qualitative accuracy hypothesis. As I specify in the book, a primitive property is (i) one whose entire qualitative nature or essence is revealed in our sensory or introspective representation of it, and thus is not identical to a property with a qualitative nature distinct from what is revealed by the representation, and (ii) one that is metaphysically simple and thus not constituted by a plurality

[12] Alvin Plantinga, "Epistemic Probability and Evil," *The Evidential Argument from Evil*, Daniel Howard-Snyder, ed., (Bloomington: Indiana University Press, 1996), pp. 69-96, at pp. 87-89.

of other properties. Properties can also be *represented as primitive*. For the redness of a sunset to be represented as primitive requires that it be represented as having that familiar simple qualitative nature revealed in visual experience of red things under normal conditions, and in such a way that excludes its being identical with any property, such as *being spectral reflectance profile S* or *being molecular basis M of spectral reflectance profile S*, whose qualitative nature is not revealed in that sensory experience.[13] It's open, I think, that how we introspect phenomenal properties generates a strong but mistaken tendency to believe that they are primitive, and that this false belief plays an important role in fueling anti-physicalist intuitions. Nonetheless, I think that the force of the knowledge and conceivability is retained given only the weaker qualitative accuracy hypothesis.

Stoljar also objects that no account of introspection is available supposing the truth qualitative inaccuracy hypothesis. In the book, I endorse Sydney Shoemaker's broad perceptual model. On Stoljar's reading of that model, introspective representations are beliefs, and in his view it's hard to see how it could be an open possibility that I am not itchy but merely believe that I am. First of all, it does turn out to be open to me to elaborate the qualitative inaccuracy hypothesis on a view in which introspective representations are beliefs, for as I develop the hypothesis, when I introspect my itchiness I come to believe that it has a qualitative nature that it in fact lacks.[14] The qualitative inaccuracy hypothesis is not eliminativist about the phenomenal

[13] Alex Byrne and David Hilbert, "Color Primitivism," *Erkenntnis* 66 (2007), pp. 73–105; David Chalmers, "Perception and the Fall from Eden," in *Perceptual Experience*, ed. Tamar Szabó and John Hawthorne (Oxford: Oxford University Press, 2006), pp. 49–125.

[14] Sydney Shoemaker, "Self Knowledge and Inner Sense," *Philosophy and Phenomenological Research* 54 (1994), pp. 249-314, at pp. 253-54. The view I spell out is

property of itchiness. On that hypothesis, it only turns out that itchiness does not have the qualitative nature that it's introspectively represented as having, not that no state of mine is an itchy one. In the Lockean view, colors lack the qualitative natures they are perceived as having, but objects are nevertheless colored. The open possibility I envision for our introspective representations of phenomenal properties is analogous on both counts.

Stoljar indicates (in correspondence) that his real concern here is that what it's like to be itchy is implausibly what it's like to merely believe that one is itchy, while on the broad perceptual model introspective representations are exclusively beliefs. In my view, the phenomenal property instance *what's it's like to be itchy* is independent of my introspective representation of it, as the broad perceptual model requires. But might my introspective representation of this phenomenal property be a belief? I'm not convinced that Stoljar is right to say that this is implausible. This introspective belief would feature a phenomenal concept, which might well account for how the phenomenal property is introspectively represented. But furthermore, in the view I prefer, introspective representations can also be perceptions and not merely beliefs, and our introspective representation of what it's like to be itchy is most fundamentally a perceptual matter. As I read Shoemaker, this view is consistent with the broad perceptual model, and so I would disagree with Stoljar on this issue of classification.[15]

---

compatible with Shoemaker's condition (7), causal production of introspective representation and (8), perception and belief-independence of what is introspected and causes the introspective representation, which constitute the central features the broad perceptual model (p. 269).

[15] That introspective representations might be perceptions and not merely beliefs is consistent with Shoemaker's conditions (7) and (8); (see note 14). Allowing that introspective representations be perceptions would not commit one to the single relevant alternative in

Stoljar's clarificatory remarks about his epistemic strategy, according to which ignorance of the physical facts explains the sense that zombies are conceivable, are sharp and illuminating.[16] As he points out, I argue in the book that the qualitative inaccuracy hypothesis should be counted as a welcome supplement to the epistemic strategy. Let me defend this claim in a different way. A general concern for Stoljar's epistemic strategy if new physical truths are added to the physical base, 'PT and ~ Q' will continue to be ideally, positively, and primarily conceivable so long the new physical truths are of the same sort that were already present in the base. To engage with the core concerns of the conceivability argument some more radical move needs to be made. My physicalist-friendly Russellian monism does so by adding physical truths of quite a different sort – i.e., truths about absolutely intrinsic properties – to the

---

Shoemaker's classification system, the act-object model, and all of its specifications (1)-(6) (pp. 252-53), in particular to the claim that what is introspected is object-like. Thanks to Daniel Stoljar and Brie Gertler for discussion of this issue.

Shoemaker's "self-blindness" objection to the broad perceptual model ("Self Knowledge and Inner Sense," pp. 271-90) is widely discussed. For replies, see William Lycan, *Consciousness and Experience*, Cambridge MA: MIT Press, 1996; Alex Byrne, "Introspection," *Philosophical Topics* 33 (2005), pp. 79-104, and Brie Gertler, *Self-Knowledge* (London: Routledge, 2010), pp. 153-59. I'm attracted to Lycan's theory of introspection, which Gertler classifies as an inner sense theory, contrasting with the acquaintance and rationality theories.

[16] Daniel Stoljar, *Ignorance and Imagination*: *The Epistemic Origin of the Problem of Consciousness* (New York: Oxford University Press, 2006).

relational/dispositional truths already there. The qualitative inaccuracy strategy's radical move is to propose that the putative qualitative natures of phenomenal properties that give rise to the conceivability of 'PT and ~ Q' are incorrectly attributed to them. If this amounted to a standard sort of eliminativism, on which the phenomena to be explained are simply denied, this move would fail to address the core concerns. But this strategy explains inaccuracy as resulting from the kind of misperception familiar from theories of color perception, so this strategy does not simply deny what is to be explained.

The positive proposal that Adams sets out in his commentary is a version of Russellian monism. Like Galen Strawson's view, it's a kind of micropsychism. Adams's proposal begins with the claim that, as his ninth grade science teacher Miss Quinn pointed out, science tells us what electricity does, not what it is, and this claim can be generalized for all physical entities. More specifically, science does not tell us about the absolutely intrinsic properties of the entities in its purview, so to rule out the physicality of micropsychic absolutely intrinsic properties on the ground that they would be fundamentally mental is questionable. It is true there is an important debate in the history of philosophy about whether there are fundamental mental entities, but if the physical is really an ontological and not a sociological category, we don't have good reason to believe that fundamental mental entities cannot count as physical. Adams raises the objection that phenomenal properties don't seem spatially extended, but responds that this may just be an artifact of how phenomenal properties are apprehended introspectively.

Perhaps there is a reason to be skeptical about the prospects of the kind of micropsychism Adams proposes. Imagine first that 'M' is the complete micropsychist truth -- it details all the phenomenal absolutely intrinsic properties of fundamental physical entities, specifying the qualitative natures of those properties. Suppose 'E' is the complete relational/dispositional truth

about electrons. Would 'MT and ~ E' be ideally, positively, and primarily conceivable? One might wonder whether there is any less reason to think 'MT and ~ E' is ideally, positively, and primarily conceivable than there is to believe that 'PT and ~ Q' is. Imagine that every fundamental particle has some absolutely intrinsic phenomenal property, and that ordinary introspectible phenomenal entities are composed of many fundamental particles of this sort. It seems as easy to conceive of any such array of fundamental micropsychic properties without electrons attracting protons as it is to conceive of any arrangement of relationally/dispositionally characterized fundamental physical particles absent some phenomenal truth.

Building on a point made by Karen Bennett, it appears that the envisioned sort of phenomenal micropsychism would need to posit fundamental or brute laws linking the micropsychist absolutely intrinsic properties with the microphysical properties they underlie, without which the truths about the microphysical properties cannot be derived from the micropsychist truths.[17] This is a reason to think that phenomenal micropsychism cannot provide a deeply illuminating explanation of the properties specified by current microphysics – any such explanation would rely crucially on brute laws.

Adams proposes a solution to this problem. As an objection to his version of Russellian Monism, he cites the concern that we are acquainted with phenomenal qualitative natures as they are in themselves, but phenomenal character is not self-presented as having the dispositional and causal relational properties discovered by physical research about electricity. More specifically, "it is scientifically important that electrical charges and discharges have quantity that can be measured, but that does not seem to be true about phenomenal qualities." The Kantian solution he suggests is that phenomenal qualities differ in intensity, and that different intensities in

[17] Karen Bennett, "Why I Am Not a Dualist," ms.

sensations correspond to different physical extensive magnitudes. However, if the correspondence amounts only to correlation, then it would seem that the ideal, positive, and primary conceivability of 'MT and ~ E' remains in place. An alternative is to advocate a qualitative inaccuracy hypothesis about the physical magnitudes, as Leibniz did, and suggest that what we perceive as extensive physical magnitudes are really phenomenal intensive magnitudes. The idealist sort of Russellian Monism that results may well be as plausible as the physicalist one I set out, at least considered independently of the inductive argument for physicalism from the success of physicalistic science, which, again, I don't evaluate in the book.[18]

Let me now turn to Lynne Baker's comments about my proposal for nonreductive physicalism about mental states. The first concerns my suggestion that the material coincidence condition for constitution might be spelled out mereologically, specifically in terms of the provision that entities x and y are materially coincident just in case they, at some level, are made out of the same parts, where 'are made out of the same parts' is cashed out as 'at some level, decompositions of x and y have all the same parts.' One objection is that decomposition is a reductive relation, so that if decompositions of x and y have all the same parts, then reductions of x and y have all the same parts. In response, if the term 'decomposition' does have this implication, then I would foreswear the use of this term, and specify instead just that x and y, at

[18] Adams also makes the point is that it may be that the physical is not a metaphysical but rather an epistemological category, and that this is arguably so on any view that characterizes the physical in terms of the entities over which the science of physics quantifies. Robert Howell discusses this problem in the first part of *Subjective Physicalism* (New York: Oxford University Press, 2013), and in response he proposes a metaphysical account of the physical inspired by Descartes's characterization of the physical as the spatial.

some level, have the same parts. The core idea is that there is a way of dividing x into parts and dividing y into parts, so that the parts of x are identical to the parts of y. The alternative specification is consistent with neither x nor y being reducible to the parts in question.

Baker also objects that such a formulation of material coincidence is at odds with a denial of reductionism, for the reason that on the nonreductivist conception any parts of a mental state M1 will be intentional, while the parts of the constituting neural token N1 won't be, so if at some level M1's and N1's parts are identical, nonreductivism will be compromised. I would deny that nonreductivism entails that all the parts of a mental/intentional state are intentional. By analogy, nonreductivism about the biological allows that parts of biological entities are non-biological, and instead just chemical. In the view I propose, a token mental state will essentially instantiate a higher-level physical compositional property, and this instantiation will have parts that are just chemical. This proposal is compatible with multiple realizability, which is what provides the impetus for nonreductivism more generally.

My schema for constitution specifies that multiple realizability is required for constitution. Baker cites a potential counterexample. I illustrate the notion of constitution with the example of a diamond being constituted by a particular lattice of carbon atoms. The concern raised is that if the lattice were different, the diamond would also be different. There are two options for defending my stance. In reply, I would argue that the conditions for individuating the lattice are distinct from the conditions for individuating the diamond. If, over time, 5% of the lattice that constitutes the Hope Diamond were to wear off, the original lattice would no longer exist, but the Hope Diamond would. But if instead it turns out to be plausible that the conditions for individuating the diamond and the lattice were relevantly the same, then, as Baker suggests, it is perhaps also plausible that the diamond is identical to but not constituted of the diamond. She

argues that there is a problem with this, since in her tire example, tire T was not multiply realizable before synthetic rubber was invented, but the relation between T and its natural rubber realize is clearly constitution, since T becomes multiply realizable after synthetic rubber is invented. I would say in response that T is multiply realizable before synthetic rubber is invented due to the nomological possibility at that prior time that T be realized by synthetic rubber.

Baker's third objection is that it is theoretically desirable that the constituter and the constituted be the same type of entity. An object can be constituted of an object, but an object can't be constituted by a state of affairs. The object head-and-hammer – let's call it 'HH' – which can exist even if its head-part and the hammer-part are detached, would be a candidate for constituting the hammer, but *HH arranged hammer-wise* would not be, for the reason that *HH arranged hammer-wise* is a state of affairs and not an object. But if this specification is granted, it seems that constitutions won't necessitate the constituted thing, for the obtaining of HH alone will not necessitate the existence of the hammer. In response, I can accept what's Baker argues is theoretically desirable and retain upward necessitation, because I can specify that the hammer is constituted just by HH, but only when it has a hammer-wise arrangement. One way to secure this result is to adopt C2 as the canonical characterization of constitution, and let the y-favorable circumstances for which 'D' stands range over circumstances of part-arrangement, and not only over external sorts of relational contexts, such as the existence of the US Federal Reserve Bank and its laws in the case of the dollar bill.[19]

I agree with Baker that global supervenience will do as much to secure physicalism generally as the upward necessitation specified in C1 and C2 (although I accept that neither is sufficient for physicalism, since both require supplementation by a no-emergent-law condition).

[19] Thanks to Ted Sider for discussion of this issue.

But in addition to securing physicalism generally, one might also want to specify what it is that makes particular entities physical. On my proposal, an important part of this account is that they are constituted of entities over which physics quantifies, where constitution is a grounding relation conceived as necessitating the entity constituted. Baker is right that my view is closer to reductionism than hers, perhaps in part because I require upward necessitation in the way specified in C1 and C2, but more clearly because I contend that mental properties are identical to higher-level physical compositional properties. Arguably, type-type identity theorists such as Smart and Place would have been quite content with this sort of view, despite the fact that these higher-level properties are not neural, but multiply realizable at the neural level. Still, this position is nonreductivist because it does not identify mental properties with properties at any level of description more basic than the mental.

Baker is right that I'm a trickle-up theorist about mental causation. I don't accept trickle-up theories about all cases of constitution. For example, if we agree that the dollar bill is constituted of the carbon-compound molecules that are coincident with it, then its causal powers don't trickle up from its constitution. Instead these causal powers are largely a function of external relational context, which features the US Federal Reserve Bank and the laws that govern it. By contrast, the causal powers of a carbon molecule are not similarly a function of external relational context. Rather, their causal powers trickle up from the causal powers of the microphysical entities that constitute it together with relations among microphysical entities internal to the molecule. Are mental states more like dollar bills or like carbon molecules? Baker is an externalist about mental state individuation. But while the arguments for mental state externalism may show that there is a plausible system of classification for mental states that is externalist, it's doubtful that they establish that their causal powers are relevantly like those of

dollar bills. For a case of a view of that sort, we can look to Nicolas Malebranche, according to whom the causal powers of mental states (to the extent that they can be said to have them at all) are inherited from divine intention and causation.[20] Contemporary externalists about the mental have not shown that mental causal powers are robustly externalist in this sense. So I maintain that the causal powers of mental states are like those of carbon atoms, and that these causal powers trickle up from their constitutions in the way specified. I don't believe that these claims compromise the nonreductivism of my approach. I would want to distinguish sharply between nonreductivism and robust external relationism about causation, and to claim that nonreductivism about the mental does not depend on this sort of externalism.[21]

References

[20] Nicolas Malebranche, *The Search after Truth*, tr. T. M. Lennon and P. J. Olscamp, Cambridge, Cambridge University Press, 1997.

[21] Thanks to Nico Silins for comments on drafts of the précis and my replies to commentators, and to Ted Sider, Lynne Baker, Daniel Stoljar and Brie Gertler for valuable comments and discussion of specific issues raised in the replies.

Baker, Lynne R. *The Metaphysics of Everyday Life*, Cambridge: Cambridge University Press, 2007.

Bennett, Karen. "Why I Am Not a Dualist," ms.

Byrne, Alex. "Introspection," *Philosophical Topics* 33 (2005), pp. 79-104.

Byrne Alex, and David Hilbert, "Color Primitivism," *Erkenntnis* 66 (2007), pp. 73–105.

Chalmers, David. "Does Conceivability Entail Possibility?" in *Conceivability and Possibility*, ed. T. Gendler and J. Hawthorne, Oxford: Oxford University Press, 2002, pp. 145–200.

Chalmers, David. "Perception and the Fall from Eden," in *Perceptual Experience*, ed. T. Gendler and A. Hawthorne and John Hawthorne, Oxford: Oxford University Press, 2006, pp. 49–125.

Gertler, Brie. *Self-Knowledge*, London: Routledge, 2010.

Howell, Robert. *Subjective Physicalism*, New York: Oxford University Press, 2013.

Kant, Immanuel. *Critique of Pure Reason*, tr. P. Guyer and A. Wood, Cambridge: Cambridge University Press, 1987.

Leibniz, G. W. *Philosophical Papers and Letters*, ed. L. E. Loemker, Dordrecht: D. Reidel, 1969.

Malebranche, Nicolas. *The Search after Truth*, tr. T. M. Lennon and P. J. Olscamp, Cambridge, Cambridge University Press, 1997.

McGinn, Colin. "What Constitutes the Mind-Body Problem," in his *Consciousness and Its Objects*, Oxford: Oxford University Press, 2004, pp. 5-25.

Nagel, Thomas. *The View from Nowhere*, New York: Oxford University Press, 1986.

Place, U. T. “Is Consciousness a Brain Process?” *British Journal of Psychology* 47 (1956), pp. 44–50.

Plantinga, Alvin. “Epistemic Probability and Evil,” *The Evidential Argument from Evil*, D. Howard-Snyder, ed., (Bloomington IN: Indiana University Press, 1996), pp. 69-96.

Russell, Bertrand. *The Analysis of Matter*, London: Kegan Paul, 1927.

Shoemaker, Sydney. “Self Knowledge and Inner Sense,” *Philosophy and Phenomenological Research* 54 (1994), pp. 249-314.

Smart, J. J. C. “Sensations and Brain Processes,” *Philosophical Review* 68 (1959), pp. 141–56.

Stoljar, Daniel. *Ignorance and Imagination*: *The Epistemic Origin of the Problem of Consciousness*, New York: Oxford University Press, 2006.

van Cleve, James. “Inner States and Outer Relations: Kant and the Case for Monadism,” in *Doing Philosophy Historically*, ed. P. H. Hare, Buffalo: Prometheus Books, 1988, pp. 231–47.