

Dualism: how epistemic issues drive debates about the ontology of consciousness*

For *The Oxford Handbook of the Philosophy of Consciousness* (Kriegel, ed.)

Brie Gertler

You feel a tickle in your toe, you taste a tangy lemon drop, you smell coffee brewing. These experiences have a distinctive feel, a qualitative character that constitutes *what it's like* to feel a tickle or taste lemon or smell coffee. Dualism about consciousness says that this qualitative character is something over and above the physical processes associated with such experiences.

Historically, dualism was motivated by theological concerns, such as the need to explain how the soul could persist into an afterlife. But most contemporary philosophical arguments for dualism are entirely naturalistic. And they do not aim to establish the existence of immaterial *substances* such as souls; rather, they aim to show that the qualitative *properties* of conscious experience are non-physical. This chapter will deal exclusively with naturalistic property dualism.

Dualism is a metaphysical view about the nature of consciousness. But it is driven largely by epistemic concerns. Dualism's chief rival, physicalism about consciousness, is also a metaphysical view that is driven largely by epistemic concerns.

A primary goal of this chapter is to correct a widespread misunderstanding about how epistemic issues shape the debate between dualists and physicalists. According to a familiar picture, dualism is motivated by armchair reflection, and dualists accord special significance to our ways of conceptualizing consciousness and the physical. In contrast, physicalists favor empirical data over armchair reflection, and physicalism is a relatively straightforward extension of scientific theorizing. This familiar picture is inaccurate. Both dualist and physicalist arguments employ a combination of empirical data and armchair reflection; both rely on considerations stemming from how we conceptualize certain phenomena; and both aim to establish views that are compatible with scientific results but go well beyond the deliverances of empirical science. My discussion highlights these neglected epistemic parallels between dualism and physicalism.

I begin this chapter by fleshing out the distinctive commitments of dualism, in a way that illuminates the interplay of epistemic and metaphysical elements within the dualist position. Section 2 outlines two influential arguments for dualism and explains how dualists defend those arguments from key criticisms. The next two sections examine the most powerful objections to dualism. Section 3 discusses the charge that dualism is inferior to physicalism as regards the theoretical virtue of *simplicity*, and hence dualist arguments bear a special burden of proof. Section 4 discusses the worry that, given reasonable assumptions, dualists must deny that our conscious thoughts and feelings genuinely *cause* our decisions and actions. I show that each of these objections to dualism depends on substantial assumptions that cannot be empirically justified. And the objection from mental causation rests on an ambitious assumption about how we conceptualize physical phenomena. Section 5 briefly reviews how epistemic considerations inform arguments on both sides of this debate.

* I'm grateful to Uriah Kriegel and Torin Alter, who offered extremely useful comments on a previous draft of this chapter.

1. What is Dualism?

1.1 *Dualism vs. physicalism*

Dualism is the thesis that consciousness is not physical—although it may arise from physical processes, it is something over and above the physical. As mentioned earlier, contemporary naturalistic dualists favor *property* dualism. Property dualism concerns the qualitative character or feel of conscious experiences. It says that these qualitative properties of experience do not consist in, and are not necessitated by, physical properties and phenomena.

Dualism contrasts with physicalism, the thesis that everything, including consciousness, is physical (or necessitated by the physical). It is sometimes assumed that physicalism must be part of any scientific worldview. Naturalistic dualism challenges that assumption.

The dualism at issue is naturalistic in both an ontological and a methodological sense. Ontologically, naturalistic dualism regards consciousness as an aspect of the natural world. It presumably resulted from evolutionary processes, and its relation to the physical is amenable to explanation by laws of nature. Methodologically, naturalistic dualism embraces the idea that the proper way to investigate consciousness is by the use of broadly scientific methods: the acquisition, analysis, and systematization of empirical data.

Given naturalistic dualism's commitment to a scientific outlook, it can be difficult to see how, precisely, this view differs from physicalism. There is now a substantial literature on the question how best to formulate dualism and physicalism, and on "the" (defining) point at issue between these positions. The formulation of dualism I will propose seems to me promising, in that it captures the spirit of the debate and is faithful to at least the dominant positions on this topic. But contributing to the debate about the proper formulation of dualism is not my purpose here. So I do not claim that my construal is preferable to all others. And I admit that there may be some views legitimately regarded as physicalist that are compatible with the dualist thesis as I construe it; there may even be some versions of dualism that reject that thesis.

I think a useful way to approach the debate between dualism and physicalism is to consider how these competing positions view the prospects for physical science. Dualists can be just as optimistic as physicalists about the prospects for empirical science generally. In particular, dualists can allow that all of concrete reality, including consciousness, is amenable to explanation by empirical science. (I include the qualification "concrete" because the status of abstract objects is orthogonal to this debate.) But dualists will deny that an account of consciousness will be part of a specifically *physical* scientific theory.

Here is an initial, relatively abstract formulation of dualism that reflects this approach.

Dualism (initial formulation). A true and exhaustive account of consciousness is beyond the reach of physical science.

This initial formulation of dualism draws on the idea that the notion of the *physical* operative in these debates is tied to physical science. This idea is, I think, very plausible; it reflects the inclination of philosophers to defer to scientists as to the particular characteristics of ordinary physical objects and phenomena.

And a link with physical science is present in both of the basic conceptions of the physical that Stoljar (2001) identifies. On the *theory-based conception*, a property is physical if and

only if it is, or is necessitated by, “the sort of property that physical theory tells us about” (*ibid.*: 256). On the *object-based conception*, a property is physical if and only if it is, or is necessitated by, “the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents” (*ibid.*: 257).¹ Neither of these conceptions identifies a specific property (such as having spatial extension) as a marker of the physical; they both defer to physical science as to the detailed nature of the properties in its domain. In this way, both of these conceptions tie the notion of the physical to the domain of physical science: what physical science investigates, or what it (ideally) discovers.

Dualism is a metaphysical thesis. So my initial formulation is apt only if the standard defining “the reach of physical science” is metaphysical: that is, only if the domain of physical science is delineated by the metaphysical features of the items it explains or posits. I will propose a metaphysical standard in 1.3 below.

However, epistemic standards for physical science also figure prominently in this debate. Epistemic standards construe physical science in terms of certain methods of investigation. Because dualism is a metaphysical position, it is not committed to the epistemic thesis that the methods of physical science are inadequate for understanding consciousness. However, that thesis plays a leading role in some arguments for dualism. So I will discuss it before turning to a fuller explication of dualism.

1.2 *An epistemic standard for physical science: objectivity*

Physical science is often described as *objective*, by virtue of its methods of data collection and analysis. Following Goldman (1997), we can characterize objective methods as those that could (in principle) be used by different researchers investigating a single phenomenon—and, if so used, would generate the same results.² Prohibiting the use of non-objective methods may seem prudent, as it bars scientists from relying on epistemically dubious sources such as mystical visions.

But a restriction to objective methods also excludes introspection. Since introspection can be used only by a single subject—the person undergoing the introspected experience—it is an exclusively first-personal, non-objective method.³

At present, some areas of physical science, including psychology and neuroscience, make essential use of introspective reports from experimental subjects. And medical trials standardly rely on subjects’ reports, e.g. about the kind and degree of pain they’re feeling. Dualists contend that this reliance on introspection is unavoidable, since one must reflect on one’s own experience to appreciate the qualitative features of conscious experiences—e.g., what it’s like to feel a tickle or smell coffee. Any means of recognizing what someone else is feeling ultimately depends on introspection, to correlate the third-personal data with the experience in question. For example, knowing that certain kinds of neuronal activity are correlated with tickles might enable a scientist to determine, through fMRI, that a particular subject is feeling a tickle. But introspective reports are required to establish the initial correlation between the neuronal activity and those experiences. And tracking neuronal activity wouldn’t enable the scientist to comprehend what the subject was feeling, according to the dualist, unless the scientist knew from her own experience *what it’s like* to feel a tickle. Since no purely third-personal knowledge will suffice for understanding what it’s like to feel a tickle, no purely objective methodology will suffice for recognizing a tickle *as such*.

Dualists maintain that this situation is inescapable. Although developments in neuroscience may enable us to identify conscious experiences in others through fMRI or more advanced techniques, first-personal observations are required both for establishing the initial correlations to such third-personal data and for comprehending the qualitative features that make conscious experiences what they are. Jaynes expresses this point forcefully, using knowledge of the brain to represent what is available through third-personal methods:

Though we knew the connections of every tickling thread of every single axon and dendrite in every species that ever existed, together with all its neurotransmitters and how they varied in its billions of synapses of every brain that ever existed, we could still never - *not ever* - from a knowledge of the brain alone know if that brain contained a consciousness like our own. (Jaynes 1976/2000: 18)

By contrast, most physicalists maintain that we could eventually overcome this reliance on introspection. They claim that every genuine property associated with conscious experience could, in principle, be detected without reliance on introspection. Physicalists generally allow that some ways of characterizing or thinking about a given property are available only to the experiencing subject. But they generally deny that any genuine *properties* of conscious experiences are accessible in principle only through introspection, or that undergoing an experience provides for substantial knowledge about the experience that is not available through other means.⁴

Now if a comprehensive understanding of tickles or other conscious experiences required having had those experiences—or even required having had conscious experiences at all—then conscious experience would differ sharply from other phenomena. After all, one needn't undergo photosynthesis or engage in asexual reproduction or experience planetary motion in order to fully comprehend those phenomena. To deny that conscious experience is in principle accessible via third-personal methods is, in effect, to posit a disparity between conscious experience and other phenomena. On this view, conscious experience differs from other phenomena in that one must have had conscious experiences in order to fully understand them.

This methodological issue exposes a fault line in this debate, concerning whether a true and exhaustive account of consciousness requires the use of first-personal methods. Most physicalists reject this idea. Dualists accept it. The corresponding claim is as follows.

The subjectivity of consciousness. A true and exhaustive account of consciousness is beyond the reach of a purely objective science.

This claim interprets the initial dualist thesis in epistemic terms: physical science cannot fully account for consciousness because its domain is limited to what is accessible via objective methods.

As an epistemic thesis, the subjectivity of consciousness does not entail, and is not entailed by, the metaphysical thesis of dualism. Still, this epistemic thesis contributes to some arguments for dualism, and it is difficult to reconcile with physicalism. Physicalists can deny the subjectivity of consciousness. Or they can argue that the presence of an epistemic divide between the physical (which is accessible through objective methods) and consciousness (which is not), does not imply that there is a metaphysical divide between them. Each of these strategies faces difficulties.⁵

1.3 *A metaphysical standard for physical science: structure and dynamics*

On my formulation, the dualist thesis is that an exhaustive account of consciousness is beyond the reach of physical science. The subjectivity of consciousness is an epistemic gloss of this thesis: it construes it in methodological terms. But to express dualism we need a metaphysical standard, one that construes physical science in terms of the kinds of things (objects, properties, and phenomena) it recognizes.

The metaphysical standard I have in mind derives from what is arguably the prevailing conception of physical science. On this conception, phenomena at higher “levels”, such as biological or astronomical phenomena, are explained by identifying phenomena at lower “levels”, such as the phenomena studied in chemistry or physics, that constitute or necessitate them. This conception is driven by the idea that all physical phenomena are ultimately constituted (or necessitated) by fundamental phenomena, falling within the domain of physics. This means that phenomena at higher levels cannot vary independently of fundamental phenomena. To put it metaphorically: once God set in place everything that falls within the domain of physics—all of the most basic entities, with their particular properties and at their particular locations, and all of the laws governing their behavior—nothing more was needed to bring it about that photosynthesis and asexual reproduction occurred, or to set the patterns of planetary motion.

There is no consensus as to what types of entities and phenomena ultimately constitute, and explain, all physical phenomena. The fundamental level may include fermions, bosons, and other “elementary” particles, with properties of mass, charge, and spin. Or, if string theory is correct, such particles may be ultimately constituted by strings whose vibrations play the roles we assign to mass, charge, and spin. Alternatively, the fundamental level may be composed of things as yet unimagined. What is important is the *form* that physical science explanations take. Physical theory explains higher-level phenomena, such as biological and astronomical phenomena, by reference to lower-level phenomena, such as those studied in chemistry and physics, that constitute or necessitate the higher-level phenomena.

Crucially, higher-level phenomena are explained by reference to the *structure and dynamics* of underlying entities. (Throughout this chapter, my discussion of structure and dynamics is deeply indebted to Chalmers 2002.) To get a sense of structural-dynamical explanations, consider a familiar fact: lukewarm water poured into a glass will conform to the shape of the glass, whereas an ice cube placed into a glass will retain its shape. Chemists explain these differing reactions by citing the structural and dynamical properties of H₂O molecules. The relevant structural features include the composition of those molecules (two hydrogen atoms and one oxygen atom) and their polarity (the uneven distribution of negative and positive charges within them). The relevant dynamical factors include the way intermolecular forces bind molecules with that polarity, and how temperature affects these bonds. Together, these structural and dynamical characteristics (combined with the structural-dynamic features of the molecular constituents of glass, etc.) explain why the bonds between H₂O molecules are tighter at lower temperatures and, hence, why ice cubes retain their shape whereas lukewarm water conforms to its container.

Alter provides this helpful summary:

[S]tructural-dynamic descriptions are those that are analyzable in formal, spatiotemporal, and nomic terms, where the formal is the logical and the mathematical, and the nomic is the domain of laws and causation. (Alter 2016: 795)

Note that in the explanation just given, facts about structure and dynamics are taken to *necessitate* the target phenomena. To return to our earlier metaphor: once God fixed the structural and dynamic facts, nothing more needed to be done to fix the facts about how H₂O at various temperatures reacts to being placed in a container. Those latter facts are necessitated by the structural-dynamic facts.

On the prevailing conception of physical science, physical science explains (higher-level) phenomena by giving structural-dynamic descriptions of (lower-level) phenomena that constitute or necessitate them. Moreover, physical science is restricted to explanations of this sort: the lower-level factors that explain the target phenomena are characterized purely in terms of structure and dynamics. This conception of physical science, applied to my initial, abstract formulation of dualism, yields a more specific thesis:

A true and exhaustive account of consciousness is beyond the reach of a science that explains its targets purely in terms of structure and dynamics.

Let's use the term "structural-dynamic phenomena" to refer to phenomena for which structural-dynamic descriptions are *exhaustive*: that is, they capture every aspect of these phenomena. On the conception of physical science just described, explanations identify structural-dynamic phenomena that constitute or necessitate the target phenomena. Combining these points to the thesis just given yields the following dualist thesis.

Dualism. Conscious experience is neither constituted nor necessitated by structural-dynamic phenomena.

This will be our working formulation of dualism. It takes conscious experience to be non-physical by virtue of its defining qualitative properties—those responsible for *what it's like* to feel a tickle or taste lemon (etc.). (Arguably, conscious experience just is the instantiation of those properties in a subject at a time.) As I mentioned earlier, I am not claiming that this is the only reasonable way to formulate dualism. But I believe that it captures the spirit of prominent dualist views, and of the larger debate more generally.

The dualist thesis says that conscious experience cannot be analyzed purely in terms of structure and dynamics, and is not constituted or necessitated by anything amenable to structural-dynamic analysis. Of course, the dualist can allow that consciousness *has* structural and dynamic features. Gazing at a blue mug full of steaming coffee, the internal structure of your visual experience may include shifting wispy trails of white above a static blue patch. Similarly, consciousness may be dynamically related to other phenomena: pains may arise from certain types of neuronal activity, and may cause us to say "ouch!". But these structural and dynamic features do not constitute or necessitate *what it's like* to have that visual experience or to feel pain, according to dualism. The dualist contends that conscious experience's intrinsic qualitative character is not amenable to structural-dynamic analysis.

More generally, dualists maintain that any attempt to describe conscious experience in purely structural-dynamic terms will leave out a crucial—really, *the* crucial—feature of conscious experience, namely, its qualitative character. Suppose that we had an exhaustive structural-dynamic description of a certain kind of neuronal activity. Suppose further (and less plausibly) that this kind of neuronal activity was perfectly correlated with pain. According to the dualist, the structural and dynamic features of this neuronal activity would not explain the occurrence of a conscious experience with the qualitative character of *pain*. In Levine’s terms, there remains an “explanatory gap” between these structural-dynamic phenomena and consciousness: we can still wonder about *why* pain arises when, and only when, that neuronal activity occurs (Levine 1983). This contrasts with the water example. When we understand the structural and dynamic features of H₂O molecules (and the structural-dynamic features of the molecular constituents of glass, etc.), we no longer wonder why lukewarm water conforms to the shape of the glass.

Dualists draw support for their view from this contrast. The behavior of H₂O is fully explained by underlying structure and dynamics, precisely because we can see how fixing the structural-dynamic facts fixes the facts about how H₂O behaves. In other words, the structural-dynamic facts *necessitate* the facts about how H₂O behaves. But we cannot see how fixing the structural and dynamic characteristics of lower-level phenomena—whether these are neurophysiological or involve fundamental entities treated by physics—fixes the facts about conscious experience. According to the dualist, this suggests that structural-dynamic facts do not necessitate the facts about conscious experience, and hence that consciousness is not physical.

Dualists deny that physical facts *necessitate* the facts about consciousness: that is, they maintain that consciousness *could* vary independently of structure and dynamics. But naturalistic dualists generally allow that, because of contingent laws of nature linking consciousness to structure and dynamics, consciousness will not *actually* vary independently of structural-dynamic phenomena. Dualism is compatible with the idea that conscious experience arises from structural-dynamic phenomena in a lawlike way.

An analogy will illustrate why, compatibly with dualism, conscious experience may arise from structural-dynamic phenomena in a lawlike way. Suppose that there exist non-physical ghostly spirits, whose presence and behavior is not constituted or necessitated by structural-dynamic phenomena. Suppose further that these spirits can be conjured by conducting a *séance*: lighting candles while reciting certain incantations, say. We can stipulate that these physical activities give rise, in a lawlike way, to the spirits’ presence. This latter stipulation is compatible with the spirits being non-physical. If the spirits are non-physical—if they have features not amenable to structural-dynamic explanation—then the laws relating physical phenomena to the spirit world are not those (the “dynamics”) governing interactions among physical entities. In this sense, they are not physical (structural-dynamic) laws.

Astronomical, biological, and chemical phenomena are necessitated by structural-dynamic phenomena at lower levels. But conscious experience is not so necessitated, according to the dualist. This brings out the sense in which dualism is *dualist*. Physical phenomena—phenomena within the domain of physical science—are structural-dynamic phenomena (or at least necessitated by structural-dynamic phenomena). But according to dualism, consciousness is neither a structural-dynamic phenomenon nor necessitated by structural-dynamic phenomena.

1.4 *The epistemic, the metaphysical, and brute necessities*

The subjectivity of consciousness thesis (section 1.2) is epistemic: it concerns what is required for recognizing conscious experience and grasping its qualitative features. Dualism is a metaphysical thesis: it concerns the nature of consciousness. Still, it might seem that the subjectivity of consciousness implies the truth of dualism. For suppose that dualism was false, and tickle experiences could be exhaustively characterized in terms of structure and dynamics (presumably the structural-dynamic features of neuronal activity or phenomena underlying such activity). In that case, we might expect objective methods to be adequate for understanding and identifying structural-dynamic phenomena. A well-informed investigator could completely understand tickles by grasping those structural-dynamic facts, and could recognize a tickle by recognizing the corresponding structural-dynamic phenomena. So if tickles were simply a matter of structure and dynamics, we would not need to rely on introspection to recognize tickles or to understand how they feel.

This is why dualism seems to follow from the subjectivity of consciousness. If dualism were false, and the qualitative features of conscious experiences were a matter of structure and dynamics, then we should expect that those features would be accessible through objective methods. So if the feel of a tickle is not objectively accessible, this suggests that dualism is true.

However, the subjectivity of consciousness—an epistemic thesis—does not imply the metaphysical thesis of dualism. To say that structural-dynamic phenomena necessitate consciousness is not to say that we can *make sense of* how or why they do so. The necessitation may instead be *brute*, in that there remains an epistemic divide (or “explanatory gap”) between structural-dynamic phenomena and consciousness. So long as conscious experiences are necessitated by structural-dynamic phenomena, dualism is false. Yet consciousness might nonetheless be subjective. It might be that first-personal methods are required for detecting and comprehending conscious experience, even if conscious experience is somehow necessitated, in a way that we can’t comprehend—that is, *brutely*—by structural-dynamic phenomena within the reach of objective methods.

If the necessary link between structural-dynamic phenomena and tickles were unintelligible to us, then even a well-informed investigator couldn’t detect a tickle by relying solely on structural-dynamical evidence. For the investigator couldn’t “read off” the qualitative features from the structural-dynamical information. On the reasonable presumption that objective methods provide only information about structure and dynamics, the investigator would need to use first-person methods to detect tickles and correlate them with structural-dynamic phenomena such as neuronal activity.

It bears repeating that this situation, in which the structure and dynamics of lower-level phenomena somehow necessitate conscious experiences but this necessitation is opaque to us, differs starkly from paradigm cases of structural-dynamic explanation. Once we know the relevant structural and dynamical facts about H₂O molecules (and about the molecular constituents of glass, etc.), we fully understand why ice cubes retain their shape when placed in glass containers. The dualist will contend that this disparity is systematic, in that brute necessities are not required for structural-dynamic explanations of phenomena other than consciousness. For the dualist, this is reason for skepticism about brute necessities.

This juxtaposition of the epistemic and the metaphysical informs the dialectical structure of the debate over dualism. Dualists generally believe that consciousness is subjective, and take the epistemic disparity between consciousness (as subjective) and the physical (as objective) to point to a metaphysical disparity, and thereby to dualism. The step from the epistemic to the metaphysical is supported by the assumption that there are no brute necessities. For their part, some physicalists deny that there is an unbridgeable epistemic gap between consciousness and the physical. This position is known as *a priori physicalism* (see Ch. 17 of this volume). Other physicalists allow that there is such a gap, but deny that this epistemic gap provides reason to think that consciousness differs in kind from the physical. This latter position embraces brute necessities, and is known as *a posteriori physicalism* (see Ch. 18 of this volume).

2. Arguments for dualism

There are three leading arguments for dualism: the Knowledge Argument (Jackson 1982), the Zombie Argument (Chalmers 1996), and the Modal Argument (Kripke 1980). In each of these, epistemic premises play a crucial role in securing the metaphysical thesis of dualism. In fact, these three arguments are all variations on the theme just mentioned: there is an epistemic gap between consciousness and physical phenomena, and hence these are ontologically distinct. I will focus on the Knowledge and Zombie Arguments, which are especially accessible and illuminating.

2.1 *The Knowledge Argument*

Although others had presented similar arguments, the most famous formulation of the Knowledge Argument is Jackson's (1982). Jackson describes Mary, a brilliant neuroscientist of the future, who has spent her entire life in a room that is exclusively black, white, and shades of gray. While in the room, Mary comes to know all of the physical facts about color vision. She knows how the surface properties of objects affect light reflectance, and she knows how the human visual system processes light at various wavelengths. For example, she knows that "red" is the term for the qualitative property that normal humans experience when their retinas are struck by light with a wavelength of 620-780 nanometers. We can even suppose that she has identified perfect correlations between types of color experiences and types of neuronal activity. But she has never seen any colors.

Jackson invites us to consider what will happen when Mary leaves the room and, for the first time, sees something red (and which she knows to be red), such as a stop sign. He thinks that Mary will learn something new about color experience. She might express this new knowledge by saying, "Oh, *this* is what it's like to see red!" Jackson argues that, since Mary already knew all of the physical facts about color experience, this fact about the qualitative character of "seeing red" experiences is not a physical fact. He concludes that physicalism is false, and so dualism is true.

The subjectivity of consciousness is pivotal to this argument. Mary had access to all of the objective facts about color experience before her release: the only thing she lacked was first-person access to color experience, since she had not had experienced color herself. So if she learns what it's like to see red only when she first sees something red, this means that objective methods will not reveal the qualitative character of color experience. This implies that objective methods will not yield a full understanding of consciousness; that is, consciousness is subjective.

Jackson thus derives metaphysical dualism from the epistemic divide between what can be known via objective methods and what requires first-person experience.

Jackson does not define “physical”, but cashing out the notion of the physical in terms of structure and dynamics preserves the spirit and force of his argument. While in the room, Mary learns all of the structural-dynamic facts about color experience; we may further suppose that she has a comprehensive structural-dynamic description of the human visual system, the behavior of light, etc. But it seems she cannot know what it’s like to see red—following standard usage, let’s call such facts *phenomenal facts*—except by seeing something red. This means that that phenomenal fact is not deducible from the facts about structure and dynamics. Now if phenomenal facts were necessitated by structural-dynamic facts, then Mary’s full knowledge of the structural-dynamic facts should enable her to deduce the phenomenal facts. After all, if Mary knew the structural and dynamic facts about H₂O molecules, she could deduce that ice cubes retain their shapes in containers, even if she’d never encountered an ice cube. To maintain that the phenomenal facts are necessitated by the structural-dynamic facts, but are not deducible from them, it to embrace the brute necessities mentioned earlier. So unless there are brute necessities, the claim that Mary learns what it’s like to see red only when leaving the room implies dualism. In this way, the “no brute necessities” assumption licenses the shift from this argument’s epistemic premises to its metaphysical conclusion.

2.2 *The Zombie Argument*

The Zombie Argument, developed in its most sophisticated form by Chalmers (1996), also aims to show that phenomenal facts are not deducible from physical facts. But instead of describing the situation of an isolated neuroscientist, Chalmers uses a more direct thought experiment. He asks you to try to conceive of a creature that is identical to you in all physical respects, but that lacks consciousness—a “zombie”. Chalmers expects that you will be able to conceive of a zombie. That is, he thinks you will detect no incoherence in the idea of a molecule-for-molecule duplicate of yourself, with identical neurophysiology, but which lacks conscious experience. For example, he thinks you can conceive of your zombie twin undergoing c-fiber stimulation without experiencing the pain that you experience when your c-fibers are stimulated.⁶ That you can conceive of a zombie—that you find nothing incoherent in the idea of a creature sharing your physical features, but lacking consciousness—means that you cannot deduce, from facts about a creature’s physical constitution, that conscious experience is or is not present. As in the Knowledge Argument, this non-deducibility is taken to show that phenomenal facts are not necessitated by physical facts.

Our formulation of dualism is well-suited to this argument. (This is unsurprising, since that formulation is taken from Chalmers’ work.) In conceiving of a zombie, one conceives of a creature that is identical to oneself in all structural-dynamic respects but lacks conscious experiences. This means that facts about conscious experience are not deducible from facts about structure and dynamics. After all, you can equally conceive that a creature exactly similar to a zombie, in all structural-dynamic respects, *does* have conscious experiences. For *you* are such a creature. Since you can’t distinguish yourself from your zombie twin *purely* on the basis of structural-dynamic features, you cannot deduce that consciousness is (or is not) present from knowledge of those features. As with the Knowledge Argument, the step from this epistemic premise to the dualist conclusion relies on the “no brute necessities” assumption. That is, it

relies on the assumption that, if the phenomenal facts were necessitated by structural-dynamic facts, then we could deduce the former from the latter.

It's worth noting that, in contrast with the Knowledge Argument, the *Zombie Argument* need not rely on the subjectivity of consciousness. Each of these arguments attempts to establish dualism by pointing to an epistemic divide between consciousness and the physical. But in the *Zombie Argument*, this divide need not be specifically a divide between the subjective and the objective. Consider the point made in the previous paragraph: you can't distinguish yourself from your *zombie twin purely* on the basis of structural-dynamic features. So facts about consciousness are not deducible from those features. Now it may well be that the best diagnosis of this non-deducibility is that consciousness is subjective whereas the physical is objective. But the *Zombie Argument* does not strictly require that the epistemic divide is a divide between the subjective and the objective.

We believe that other human beings have conscious experiences. And arguably, we infer that they are conscious from their physical features and behavior: after all, we don't have anything else to go on. (Particularly salient here is the capacity for articulate speech, which zombies also exhibit.) However, the inference from others' physical similarity to the conclusion that they are conscious is not deductive. You can imagine, in a twist on a classic horror film trope, that the human-like creatures around you are zombies, with no conscious inner life. (This scenario is at odds with most versions of naturalistic dualism, which take consciousness to arise from the physical contingently but in a way governed by laws of nature; see note 6.) That others are physically similar to you is strong but not absolutely conclusive evidence that they have conscious experiences roughly similar to your own.

2.3 *Can we really perform these thought experiments?*

Some physicalists charge that these thought experiments are unreliable, as we are not equipped to competently evaluate the scenarios at issue. As applied to the Knowledge Argument, the worry is that even a well-informed scientist doesn't have the information needed to meaningfully imagine knowing all of the physical (structural-dynamic) facts about color experience. So we are not in a position to determine whether such knowledge would enable us to deduce the phenomenal facts. As applied to the *Zombie Argument*, the worry is that even a well-informed scientist doesn't have the information needed to truly conceive of a perfect physical (structural-dynamic) duplicate, and hence we cannot evaluate whether the possibility of a zombie is coherent.

It's true that neuroscience is still in its early days, and theories in fundamental physics remain speculative. In engaging in these thought experiments we must therefore rely on a general conception of what it is to be *physical*. Following Chalmers (2002), I have suggested that we conceive of the physical as what can be exhaustively characterized in terms of structure and dynamics. But this suggestion does not silence the current worry. For we don't know what kinds of structural-dynamic phenomena are yet to be discovered, or how our models of physical phenomena might evolve (while remaining structural-dynamic). So we cannot conceive of all of Mary's knowledge, or all of the properties that would be shared by our structural-dynamic duplicate, in detail.⁷

The dualist will maintain that the details aren't needed to perform the thought experiment, as careful reflection on what it's like to feel a tickle or smell coffee enables us to

recognize that such qualitative features of experience are not amenable to any structural-dynamic explanation, regardless of the specific structural-dynamic details. This response highlights the significance of conceptual matters in this debate. The dualist arguments rely on the accuracy of our phenomenal concepts, developed through introspective reflection on experience, and also draw on our physical concepts. As we will see below, conceptual matters are also central to physicalist objections to dualism. These objections are grounded in confidence about our ways of conceptualizing the physical, as expressed in our physical theories. They presume that our conceptualizations are not only useful but represent physical reality accurately and comprehensively.

Perhaps most importantly, our ignorance about the neuroscience of consciousness and the nature of fundamental physics cuts both ways. For the greater our ignorance about these matters, the less confident we should be about what form a comprehensive theory of concrete phenomena will take. In particular, recognizing that advances in fundamental physics may alter our current models of physical phenomena should limit our confidence that all of concrete reality is amenable to explanation in structural-dynamic terms—or perhaps even in any terms that qualify the successful theory as a *physical* theory, in a non-trivial sense.

2.4 *The use of armchair reasoning*

The Knowledge and Zombie arguments rest on the use of introspection and *a priori* reflection. This exposes them to doubts about whether such “armchair reasoning” is legitimate as a means of discerning metaphysical truths about concrete reality.

The worry is that armchair reasoning seems to reveal only *epistemic* facts. We may discover, through introspective reflection on the feel of a tickle, that phenomenal properties appear not to be a matter of structure and dynamics. Armchair reasoning may also reveal that we cannot deduce phenomenal facts from physical facts, and that we can conceive of zombies. But these are merely epistemic matters, insufficient to establish dualism.

The relation between the epistemic and the metaphysical is a philosophically foundational one. Positions on this issue are grounded in complex accounts of thought, reference, and modality. Fortunately, we needn't go into the details of those accounts to determine whether, in the context of this debate, it is legitimate for the dualist to use armchair reasoning. This is because the case for physicalism equally rests on armchair reasoning. Let me explain.

According to physicalism, facts about structure and dynamics *necessitate* all facts (about concrete reality—I will omit this in what follows). That is, it is impossible for any facts, including facts about conscious experience, to vary independently of facts about structure and dynamics. Now in order to establish that something is necessary or impossible, armchair reasoning is indispensable. Ordinary observation can establish that something is *possible*: if you see a cat on the mat, you can safely conclude that it is possible for cats to be on mats. But claims about what is necessary or what is impossible cannot be established through observation.

For example, the discovery that every creature undergoing a certain kind of neuronal activity is also experiencing pain would not show that these were linked with metaphysical necessity. This might be simply a coincidence. More plausibly, they may be linked nomologically: laws of nature may ensure that pain arises from neuronal activity of that kind. And if the link

between consciousness and the physical is only a matter of contingent laws of nature, then dualism is true (as our séance example illustrated). So nothing that could be established simply through empirical observation—not even perfect correlations between types of neuronal activity and types of conscious states—could resolve this debate in favor of physicalism.⁸ The case for physicalism requires going beyond empirical observations (which concern what’s actual) and generalizations from those observations. The dualist can allow that, because of contingent but lawlike regularities, conscious experience *will not* vary independently of the underlying physical phenomena. But the physicalist thesis requires more than this. It requires that conscious experience *could not* vary independently of the physical.

As the next two sections will illustrate, the case for thinking that consciousness is *necessitated* by the physical relies on armchair reasoning. Specifically, it relies on significant assumptions about the nature of the physical, assumptions that cannot be verified by empirical science or reasonably inferred from its findings.

3. Dualism and Theoretical Simplicity

We now turn to the first influential challenge to dualism. This challenge stems from the idea that, given the success of physical science generally, a physicalist theory of consciousness will provide for a theory of concrete phenomena that is simpler than dualism, and hence possesses greater explanatory power.

The physical sciences are rightly regarded as exemplars of knowledge production, given their remarkable success in explaining and predicting a wide range of phenomena. So it is natural to hope that they will someday provide a comprehensive account of all concrete phenomena, including consciousness.

If that hope were realized, the resulting theory would have the virtue of *simplicity*. Suppose that all phenomena, including consciousness, were amenable to explanation in terms of structure and dynamics. In that case, a single explanatory framework could be adequate for explaining all phenomena. The resulting physical theory might well be simpler than dualism in two ways. First, it might include fewer basic kinds of things in its ontology, since all of the basic entities it recognizes will be amenable to structural-dynamic analysis. (This dimension of simplicity is known as *parsimony*.) Second, it might include fewer fundamental laws, since it need not include special fundamental laws linking structural-dynamic phenomena to consciousness. (This is known as *elegance*.)⁹

Simplicity concerns exert a powerful, often covert influence on the dualism-physicalism debate. I suspect that it is the appeal of greater simplicity that is responsible for the sense that dualism faces a greater initial burden of proof.

Whether simplicity considerations provide good reason to favor physicalism depends on how the arguments for dualism fare. In general, simplicity concerns guide theory choice only when the theories being compared accommodate the data equally well. And the arguments for dualism attempt to show that physicalist theories don’t accommodate all of the data about conscious experience. The thought experiments in the Knowledge Argument and the Zombie Argument are designed to lead us to conclude that our own introspective data—data we glean from introspecting our own experiences—are not accommodated by a physicalist theory. In our

formulation, these data are not accommodated by a theory that explains phenomena purely in terms of structure and dynamics.

In response, the physicalist can argue that we should be wary of the claim that these data cannot be explained by a physicalist theory, precisely because respecting that claim requires sacrificing simplicity. After all, the physicalist might say, the importance of simplicity raises the bar for arguments aiming to establish dualism. And this high bar cannot be cleared simply by observing that some data about experiences appear to resist structural-dynamic explanation.

To assess this issue, we need to understand the importance of simplicity. How much does simplicity matter, in evaluating a theory? And *why* does it matter? These questions are surprisingly difficult to answer. While simpler theories are generally preferred in the sciences, the preference for simplicity does not translate easily to philosophy. The scientific preference may be driven by instrumental or even aesthetic considerations, whereas the philosopher's exclusive concern is with *truth*. And there is no clear reason to think that a simpler theory is more likely to be true.

A natural thought is that “simpler theories are more likely to be true” might be established inductively. An inductive argument could cite the fact that simplicity considerations guide theory choice in the sciences, and so take the predictive success of the sciences as a basis for inferring that greater simplicity indicates a greater likelihood of truth. But as Huemer (2009) demonstrates, this strategy is problematically circular. We regard scientific theories as true because the truth of their claims is the *simplest* explanation of their predictive success. The inference from predictive success to truth depends on the assumption that simplicity is linked with truth, and so it cannot (non-circularly) establish that link.

The prospects for an empirical justification for favoring simpler theories, in the search for truth, seem dim. Sober argues that, while favoring simpler theories may be justified as a means of maximizing predictive success, there may be no way of justifying the assumption that a simpler theory is more likely to be true (Sober 2009).¹⁰

The first challenge to dualism is physical science's record of success, and the corresponding promise of a unified physicalist theory encompassing consciousness and other phenomena. But whether a unified physicalist theory is possible depends on how the dualist arguments fare. Some contributors to this debate take physicalism's greater simplicity to warrant skepticism about data that threaten this unification project—e.g., introspective data that supports the Knowledge and Zombie arguments. This strategy invests simplicity with an importance and relevance to truth that may be impossible to justify—and seems not to be open to empirical justification. So we should be wary of taking the perceived threat to simplicity as grounds for skepticism about the data used to support dualism.

4. Dualism and Mental Causation

The most influential objection to dualism—and, correspondingly, the most influential argument for physicalism—concerns mental causation. The problem of mental causation is standardly regarded as the most serious challenge to dualism. This challenge takes many forms. My focus will be on the form that has dominated discussion for the past 30 years or so, namely, the exclusion argument.

4.1 *The exclusion argument*

We ordinarily assume that conscious experiences have physical effects. Ticks cause us to laugh; pain causes us to seek analgesics. But a powerful argument suggests that, if conscious experiences are something over and above the physical, then they do not cause physical events. According to this argument, dualism implies that conscious experience is *epiphenomenal*: it may be caused by physical events or processes, but it does not have physical effects. Some dualists embrace epiphenomenalism about consciousness, while others argue that conscious experience, although non-physical, can causally affect the physical realm. The problem of mental causation is widely regarded, by physicalists and dualists alike, as the most serious challenge facing dualism.

Why think that dualism implies that conscious experience has no physical effects? In a nutshell, the worry is this. In order to explain a physical event, we never *need* to invoke anything non-physical—or so it seems. For example, we may one day be able to fully explain how ticks cause laughter and how pain causes us to say “ouch!” and to seek analgesics. Such explanations would presumably invoke only *physical* causes: in our terms, only phenomena that are amenable to structural-dynamic explanation. (For brevity, I will continue to use “physical”.) Given the availability of purely physical explanations of human behavior, no explanatory work remains for non-physical factors to perform. So if the qualitative character of ticks and pains—how these sensations *feel*—is non-physical, then reference to this qualitative character is not needed to explain laughter or analgesic-seeking behavior. More generally: if, for every physical event that can be causally explained, a causal explanation in purely physical terms is available (in principle), then the non-physical seems inconsequential, as regards causal explanations of physical phenomena.

This problem, which is known as the *exclusion problem*, was developed in its contemporary form by Jaegwon Kim (1989a). The availability of purely physical explanations seems to leave no work to be done by explanations citing non-physical factors, and hence to *exclude* the latter.

Now even if non-physical features of conscious experiences are not *needed* for causally explaining physical events, this does not mean that those features are causally inert. Imagine that you’re awoken by two simultaneous alarms. Each of these alarms is redundant, in that the presence of the other alarm means that neither one is *needed* for the task of awakening you. But adding the second alarm didn’t *silence* the first one; neither alarm is robbed of its causal power by the presence of the other. Similarly, a causal explanation of laughter may invoke the (non-physical) *feel* of a tickle even if, given the availability of a physical explanation, the explanation that invokes that non-physical feel is redundant.

Suppose that, for every physical event for which a causal explanation is available, there is an explanation in purely physical terms. Non-physical properties of experience might nonetheless be invoked in (admittedly redundant) causal explanations of physical events. In that case, every physical effect (such as saying “ouch!”) of a non-physical conscious experience (such as feeling a pain) is *overdetermined*: there is a physical cause, capable of bringing about the effect all on its own (perhaps neuronal activity), and an additional non-physical cause (the pain sensation), also capable of bringing about the effect all on its own. The specter of rampant overdetermination—a pervasive redundancy in nature—is unappealing.

The dualist has three unattractive options. The first is to deny that conscious experiences have causal efficacy—that is, to embrace epiphenomenalism. The second is to maintain that

conscious experiences are causally efficacious, by allowing the type of rampant overdetermination just described. The third option is to deny that every physical effect could be explained purely by reference to physical phenomena.

The physicalist can avoid this problem by saying that conscious experiences *just are* physical events (so qualitative character is a physical feature).¹¹ In that case, conscious experiences and physical events don't compete for the title "cause", since they are one and the same. So physical explanations don't *exclude* explanations that cite conscious experiences.

Papineau (2002) provides a tidy formulation of this argument.¹²

- (1) Some conscious experiences have physical effects. (This is the rejection of epiphenomenalism.)
- (2) All physical effects are fully caused by purely *physical* prior histories. (That is: for every physical event that can be causally explained, a purely physical explanation is in principle available.)
- (3) The physical effects of conscious experiences aren't always overdetermined by distinct causes.

Conclusion: The conscious experiences mentioned in (1) must be identical with some part of the physical causes mentioned in (2).

Suppose we want to identify what caused your laughter. To secure a non-superfluous explanatory role for the feel of the tickle, reference to that feeling should be essential to the causal explanation. This requirement is easily met by simply *identifying* that feeling with a type of physical state. On this physicalist proposal, "tickle sensation" and "neuronal activity *N*" refer to the same thing—they are two terms for a single type of state. To say that that state caused your laughter is to provide a physical explanation (consonant with premise 2) that invokes conscious experience (confirming premise 1) but avoids overdetermination (confirming premise 3).

The physicalist thesis that avoids the exclusion problem, and that serves as the conclusion of this argument, is especially strong. The idea that tickle sensations are strictly identical to a certain kind of physical event implies that tickle sensations cannot be realized in other physical types—other types of brains, with different structural-dynamic characteristics. For this reason, most physicalists deny that conscious experiences are strictly identical to physical events. Most favor the more moderate physicalist thesis that conscious experiences are necessitated by, but not identical to, physical events.

Moderate physicalism also faces the exclusion problem (Kim 1989b). If physical effects can be causally explained purely by reference to physical causes, and physical causes are not identical to conscious experiences, then physical explanations exclude explanations invoking conscious states. (See Bennett 2007 for an illuminating discussion of this problem and of the responses available to the moderate physicalist.)

4.2 *The completeness premise*

I believe that epiphenomenalism, the denial of premise 1 of the argument, is less damaging to our commonsense views than it initially appears (Gertler forthcoming). But my discussion here will center on premise 2, which is known as the *completeness* premise. It says that

the realm of the physical is explanatorily complete, in that we never need to invoke anything outside that realm to explain any (explainable) physical event.

Applying our conception of the physical yields the following interpretation of the completeness premise.

Structural-Dynamic Completeness:

For every effect that can be exhaustively characterized in structural-dynamic terms, an explanation invoking only structural-dynamic phenomena is (in principle) available.¹³

This statement seems plainly true. By conceptualizing an effect in structural-dynamic terms, we more or less ensure that it is amenable to structural-dynamic explanation. After all, structural-dynamic characterizations of a phenomenon include its causal and nomic properties: the kinds of interactions it enters into, and the laws governing its behavior. So to characterize an event in structural-dynamic terms is already to locate it within the explanatory framework of structure and dynamics. What this shows is that much of the weight of the argument from mental causation is borne by the way that the effects to be explained—the “physical” effects—are conceptualized.

In order for the exclusion argument to succeed, our way of conceptualizing the events to be explained—the “physical effects”—must be not only accurate but also exhaustive. The argument would fail if there were features of these target effects that were not simply a matter of structure and dynamics. For suppose that the physical effect to be explained had some such features. (These could be characteristics at the fundamental level of physics, which underlie or complement the phenomenon’s structure and dynamics.) In that case, the availability of a structural-dynamic explanation would not render other explanations superfluous. These other explanations might do needed explanatory work as regards the aspect of the phenomenon that resists structural-dynamic analysis. This would neutralize the overdetermination worry, and thereby undermine the exclusion argument.¹⁴

These reflections reveal an assumption of the exclusion argument: that our way of conceptualizing the physical, in terms of structure and dynamics, is not only accurate but also *exhaustive*, in that it captures all explanatorily salient features of physical reality. This assumption seems reasonable enough, so long as we have no reason to think that there’s a dimension of physical phenomena that resists structural-dynamic analysis or explanation. But it’s worth noting that this is a very substantial assumption. It requires an especially ambitious view about the relation between how we conceptualize reality and reality itself. The previous section discussed a moderate (but still questionable) claim about this relation, namely that if two theories—two ways of conceptualizing and explaining a set of phenomena—accommodate the data equally well, then the simpler theory is more likely to be true. If we had reason to accept that claim, it could be used to establish that the way our physical theories conceptualize (physical) reality, in terms of structure and dynamics, is accurate: that is, physical reality truly has the structural-dynamic features we attribute to it. But even this argument, which relies on a questionable claim about the link between simplicity and truth, is silent on the question whether structure and dynamics *exhaust* the physical. For it says nothing about whether there is anything *beyond* structure and dynamics.

The assumption that there is no aspect of physical phenomena that evades structural-dynamic explanation could be challenged on a variety of grounds. In fact, the worry about causal exclusion could fuel just such a challenge. Positing an aspect of physical effects that is not amenable to structural-dynamic explanations provides a way to avoid both epiphenomenalism and overdetermination.

Let me sum up. The leading form of the problem of mental causation, the exclusion argument, depends on an ambitious assumption: that our conceptualization of physical phenomena, in terms of structure and dynamics, is not only accurate but also exhaustive. This assumption goes well beyond the kind of background assumption required to avoid skepticism, viz. that our concepts correspond (at least roughly) to features of reality itself. It also goes well beyond the idea that simpler theories are more likely to be true. And it is a conceptual, philosophical claim, not a scientific one: nothing empirical could establish that physical theory *comprehensively* captures the underlying reality it (perhaps accurately) describes. So this objection to dualism, like the objection from simplicity, relies on confidence that our way of conceptualizing the physical, in physical theories, is linked to the physical itself. But the required link here is particularly robust. It is that there is no aspect of physical phenomena beyond what is captured by our theories; in other words, physical phenomena consist entirely in structure and dynamics (and what is necessitated by structure and dynamics).

5. The Epistemic Sources of Dualism and Physicalism

Arguments on both sides of this debate rely on epistemic premises: claims about how we represent, conceptualize, and learn about conscious experience and/or physical phenomena. These arguments also rely on strategies for justifying the inference from their respective epistemic premises to their respective metaphysical conclusions.

For the dualist, the epistemic premises include the subjectivity of consciousness and the (apparent) conceivability of zombies. The metaphysical conclusion is that conscious experience is not identical to or necessitated by structural-dynamic phenomena. The bridge from the epistemic to the metaphysical is the “no brute necessities” assumption: if there were *necessary* connections between conscious experience and structural-dynamic phenomena, these connections would be intelligible to us.

For the physicalist, the epistemic premises include the claim that conceptualizing physical phenomena in terms of structure and dynamics—the conceptualization embodied in physical theory—is not only accurate but exhaustive. The metaphysical conclusion is that conscious experience is amenable to explanation in these terms: it consists in, or is necessitated by, structural-dynamic phenomena. In the arguments we considered here, the bridge from the epistemic to the metaphysical depends on simplicity considerations. Suppose that all concrete phenomena other than consciousness consist in, or are necessitated by, purely structural-dynamic phenomena (this is the epistemic premise). In that case, a structural-dynamic explanation of consciousness will yield a theory of concrete reality that has a relatively high degree of theoretical simplicity. In particular, given the epistemic premise, a structural-dynamic explanation of consciousness will yield a theory that is simpler than a theory that disavows structural-dynamic explanations of consciousness—that is, a dualist theory. The bridge from the

epistemic premise to the metaphysical conclusion is the assumption that if two theories accommodate the data equally well, the simpler theory is more likely to be true.

The exclusion argument fits this pattern. The epistemic premise is that structural-dynamic explanations of non-mental phenomena are accurate and comprehensive—they leave no dimension of the target phenomenon unexplained. In light of that premise, positing alternative explanations—explanations citing phenomena that are not structural-dynamical—compromises theoretical simplicity, by requiring overdetermination. Embracing epiphenomenalism also threatens simplicity. For the causal inefficacy of conscious experience makes it a “nomological dangler”: something that (we presume) arises from physical processes but does not give rise to anything. This picture is plainly inelegant. The threat to simplicity seems to be what Smart has in mind when he complains that “the laws whereby these nomological danglers would dangle ... would be like nothing so far known in science. They have a queer ‘smell’ to them.” (Smart 1959: 142-3)

One’s position on dualism will not be determined simply by how one regards conceptual arguments, or by the extent of one’s deference to physical science and theorizing. Arguments on both sides of this debate presume that the way we conceptualize a phenomenon can justify claims about that phenomenon’s metaphysical nature. And arguments on both sides employ distinctively philosophical claims—claims outside the purview of science—about the status and proper domain of physical science. Where one stands on the question of dualism chiefly depends on where one stands on more specific epistemic questions: the epistemic premises of the arguments for and against dualism, and the way these arguments justify the inference from those premises to their respective metaphysical conclusions.

Works cited

Alter, T. (2016), ‘The Structure and Dynamics Argument against Materialism’, in *Noûs* 50: 794–815.

Alter, T. & Nagasawa, Y., eds., (2015), *Consciousness in the Physical World: Perspectives on Russellian Monism* (Oxford University Press).

Balog, K. (2012), ‘In Defense of the Phenomenal Concept Strategy’, in *Philosophy and Phenomenological Research* 84: 1-23.

Bennett, K. (2007), ‘Mental Causation’, in *Philosophy Compass* 2: 316–337.

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

Chalmers, D. J. (2002), ‘Consciousness and its Place in Nature’, in D.J. Chalmers (ed.), *Philosophy of Mind: Classic and Contemporary Readings*. Oxford University Press), 247-72.

Chalmers, D. J. (2006), ‘Phenomenal concepts and the explanatory gap’, in T. Alter and S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (Oxford University Press), 167-94.

- Dominus, S. (2011), 'Could Conjoined Twins Share a Mind?' *New York Times Magazine*, May 29, 2011. (<http://www.nytimes.com/2011/05/29/magazine/could-conjoined-twins-share-a-mind.html>)
- Gertler, B. (forthcoming), 'Acquaintance, Parsimony, and Epiphenomenalism', in S. Coleman (ed.), *The Knowledge Argument Then and Now* (Cambridge University Press).
- Goldman, A. (1997) 'Science, Publicity, and Consciousness', in *Philosophy of Science* 64: 525-45.
- Howell, R. J. (2013), *Consciousness and the Limits of Objectivity: The Case for Subjective Physicalism* (Oxford University Press).
- Huemer, M. (2009), 'When is Parsimony a Virtue?', in *Philosophical Quarterly* 59: 216-36.
- Jackson, F. (1982), 'Epiphenomenal Qualia', in *Philosophical Quarterly* 32: 127-36.
- Jaynes, J. (1976/2000), *The Origins of Consciousness in the Breakdown of the Bicameral Mind* (Houghton Mifflin).
- Kim, J. (1989a), 'Mechanism, Purpose, and Explanatory Exclusion', in *Philosophical Perspectives* 3: 77-108.
- Kim, J. (1989b), 'The Myth of Nonreductive Physicalism', in *Proceedings and Addresses of the American Philosophical Association* 63: 31-47.
- Kripke, S. (1980), *Naming and Necessity* (Harvard University Press).
- Levin, J. (2006), 'What is a phenomenal concept?', in T. Alter and S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (Oxford University Press), 87-110.
- Levine, J. (1983), 'Materialism and Qualia: the explanatory gap', in *Pacific Philosophical Quarterly* 64: 354-61.
- Papineau, D. (2002), *Thinking About Consciousness* (Oxford University Press).
- Russell, B. (1927), *The Analysis of Matter* (Kegan Paul).
- Smart, J.J.C. (1959), 'Sensations and Brain Processes', in *The Philosophical Review* 68: 141-56.
- Sober, E. (2009), 'Parsimony Arguments in Science and Philosophy—A Test Case for Naturalism', in *Proceedings and Addresses of the American Philosophical Association* 83: 117-55.
- Stoljar, D. (2001), 'Two Conceptions of the Physical', in *Philosophy and Phenomenological Research* 62: 253-81.

¹ The fact that "physical" occurs in both conceptions does not challenge their legitimacy, since they are not intended as reductive analyses.

² This is a rough characterization, adapted from Goldman’s very useful discussion of the alleged “publicity requirement” in science (Goldman 1997). Goldman considers various ways of refining this characterization.

³ The possibility that a single experience might be introspected by multiple persons is raised by some cases of craniopagus twins—twins who are fused at the cranium. A pair of craniopagus twins born in 2006, Krista and Tatiana Hogan, are said to share some sensations. When one twin is tickled both feel the tickle (Dominus 2011). Whether this challenges the claim that an experience can be introspected only by a single subject depends on further details: specifically, whether both twins have introspective access to a single conscious experience, or whether instead the tickling causes each to have her own tickle experience.

⁴ Physicalists who use the so-called “phenomenal concept strategy” to block arguments for dualism generally claim that there are ways of conceptualizing the qualitative character of experiences that are available only to those who have had such experiences. But in most versions of this strategy, these conceptualizations—so-called “phenomenal concepts”—are not required for a comprehensive understanding of the qualitative property itself (e.g. Levin 2006). Some physicalists, such as Balog (2012) and Howell (2013), allow that introspection is required for fully comprehending the qualitative character of conscious experience, but argue that this epistemic fact is compatible with a physicalist view about the nature of consciousness. Chalmers (2006) challenges this general physicalist strategy. I raise doubts about Balog’s and Howell’s versions of it in a forthcoming paper (Gertler forthcoming).

⁵ The difficulty of the former strategy is reflected in the intuitive appeal of the Jaynes passage above, and in the force of the Knowledge Argument for dualism (section 2.1). Difficulties facing the latter strategy include those faced by the second phenomenal concept strategy (see note 4).

⁶ Crucially, we are not to assume that the duplicate is subject to the same laws of nature. For as the séance example illustrated (section 1.3), dualism is compatible with the idea that consciousness arises from physical phenomena as a matter of contingent laws.

⁷ Another concern is that we may err even in applying those concepts we do possess: that is, in attempting to conceive of a zombie. Chalmers addresses this issue by appealing to an idealized notion of conceivability. As he notes, this notion idealizes away from our cognitive limitations in much the same way as the notion of *a priori* truths idealizes away from those limitations: a mathematical equation may be *a priori* even if, because of its complexity, it cannot be properly evaluated by unaided human cognition.

⁸ Since dualism is a thesis about what is possible, rather than a thesis about what is necessary or impossible, this debate could ostensibly be resolved by empirical observation: e.g., the discovery of two creatures that were exactly similar, physically, but that differed in their conscious experiences. But verifying this would be difficult (to say the least). And it would undermine the standard naturalist form of dualism, which says that there is a lawlike relation between consciousness and the physical.

⁹ That a theory is compatible with physicalism does not *entail* that it is simpler than a dualist theory. Suppose that qualitative features of experience are intrinsic to structural-dynamic entities, but are not themselves fixed by structural-dynamic factors. This entails dualism (on my formulation of dualism). Yet it is compatible with a high degree of both parsimony and elegance: e.g., a kind of monism that allows for a sparse set of basic laws and principles. Views of this kind, often inspired by Russell (1927), have recently grown in prominence. (See chapters 14 and 22 of this volume, and the papers

in Alter and Nagasawa 2015.) There is some controversy as to whether this monistic view strictly qualifies as dualism or as physicalism; such controversies mean that any formulation of these positions, including mine, will be open to question. But the fact that this view seems, to many, to be closer in spirit to dualism than to physicalism illustrates the inadequacy of the term “dualism”, as that term suggests that dualism is incompatible with monism.

¹⁰ Sober expresses skepticism about the association of parsimony with truth, as regards “the mind/body identity theory”, a version of physicalism. (The term “model selection theory” refers to a particular operational interpretation of parsimony.)

Placing the mind/body identity theory and dualism within the context of model selection theory requires one to think of the contending theories in terms of their predictive accuracy, not their truth. Metaphysicians may balk at this, proclaiming that they don’t care about predictive accuracy and want only to figure out what is true. I am not arguing against that preference. Rather, my point is that the parsimony argument for the identity theory finds a natural home in the model selection framework. If there is another treatment of the argument that establishes its connection with truth, I do not know what that treatment is. (Sober 2009, 137)

¹¹ On the plausible view that a conscious experience is the instantiation of qualitative properties in a subject at a time, the claim that conscious experiences are physical events implies that qualitative properties are physical.

¹² This formulation is paraphrased from Papineau 2002: ch. 1.

¹³ Strictly speaking, on our formulation physical events are constituted *or necessitated* by structural-dynamic phenomena. I elided this qualification for brevity. It does not change the basic point since, trivially, an effect necessitated by structural-dynamic phenomena is explainable in terms of structure and dynamics.

¹⁴ Following Russell (1927), some have suggested that structural-dynamic characterizations neglect the intrinsic or categorical features of physical things. This opens up the possibility of a variety of alternatives to classic dualism and physicalism. See note 9.