

Idealism: Putting Qualia To Work*

Michael Pelczar

1 Introduction

Metaphysical idealism is the mirror-image of physicalism about the mental: where physicalists contend that the mental facts of our world supervene on the physical facts (but not vice versa), idealists contend that the physical facts of our world supervene on the mental facts (but not vice versa).²

Like physicalism, idealism is a kind of monism. According to idealists, the fundamental features of our world (or at least its fundamental contingent features) are all of one kind—the mental kind. Unlike physicalists, however, idealists try to achieve monism without reducing consciousness to something ostensibly more basic, or identifying consciousness with something that we previously didn't realize was consciousness (like brain states).

It's been a long time since idealism was part of the mainstream philosophical conversation. To get a feel for the theory, and to see its attractions, it helps to go back to the days when most philosophers considered idealism a live option.

A good starting point is Kant, whose metaphysics, though not itself strictly idealistic, is a watershed divide between two major idealist world-views: the traditional idealism of Leibniz and Berkeley, and the phenomenalism of J.S. Mill and the sense-datum theorists of the early 20th century. §2 of this chapter presents Kantian metaphysics in its bare essentials. §3 presents

* To appear in *Oxford Handbook of the Philosophy of Consciousness*, U. Kriegel, ed.

² Here and throughout, “supervene” is used in its minimal metaphysical sense: the X features of our world supervene on its Y features just in case metaphysically possible worlds exactly like ours in terms of their Y features contain all the X features that our world contains. (Those who prefer grounding to supervenience can substitute the ensuing supervenience-talk with its closest approximation in terms of grounding.)

traditional idealism as a view that results from modifying Kantian metaphysics in one direction; §4 presents phenomenalism as a view that results from modifying Kantian metaphysics in the opposite direction. In §5, we consider some important challenges to the supervenience claim that lies at the heart of all idealist theories. §6 concludes with some brief speculation about what it would take for idealism to reverse its long-declining fortunes.³

2 Kantian Metaphysics

Beliefs about the ultimate causes of our experiences have changed dramatically over the millenia, from combinations of the Four Elements, to geometric configurations of Democritean atoms, to dynamical systems of Newtonian bodies, to excitation states of quantum fields. By contrast, beliefs about what the physical world contains have remained highly stable. The ancient Greeks, the natural philosophers of the Enlightenment, and scientists of the 21st century all agree that the world contains trees, despite having profoundly different beliefs about the nature of what gives us our experiences of trees.

A natural explanation for this agreement is that people throughout history have thought, and continue to think, that in order for there to be trees, it's enough for our world to be a place where experiences tend to occur in ways that are suggestive of trees. What has changed over the years are people's opinions about what accounts for that tendency.

This suggests a metaphysics. Our world has the power to cause experiences. For each physical state of affairs that holds in our world, our world has an experience-causing power, or

³ The correct interpretation of historic idealists' views is controversial. The interpretations I assume here are, I hope, recognizably mainstream, but I've chosen them mainly for the light they shed on idealism as a living theory, and when I choose one interpretation over another, it's without any pretense of settling the associated scholarly debate.

combination of such powers, the existence of which is metaphysically sufficient for that physical state of affairs.

David Chalmers has proposed a metaphysics along these lines. According to Chalmers, ostensibly skeptical hypotheses, such as the hypothesis that all our experiences arise from interactions between some envatted brains and a supercomputer, are really just quirky metaphysical hypotheses consistent with the truth of our everyday beliefs about the world's physical contents. Assuming that the computer has the same experience-causing powers as whatever it is that actually gives us our experiences (of trees and other things), it follows that there really are trees in the envatted brains scenario. In this view, discovering that we were, in fact, envatted brains, would be like discovering that physical objects are fundamentally constituted by protons, neutrons, and electrons, rather than Earth, Water, Air, and Fire.⁴

The best-known proponent of this style of metaphysics is Kant.⁵

Here's the Kantian picture: there are entities that broadcast signals. The only thing we can know about these entities is that they broadcast such signals. The signals don't exist in a physical form, and they don't propagate through time or space. The only things capable of receiving the signals (as far as we know) are conscious minds, like ours. But the entities that broadcast the signals do so regardless of whether there are any minds to receive them. For physical things—say, trees—to exist is for it to be the case that if there were minds receptive to all such signalling, then the signals would cause those minds to have experiences that were collectively suggestive of trees.

What does it mean, to say that a collection of experiences is “suggestive of trees”? Kant doesn't offer much detail on this, but we can think of it in terms of how the experiences in the

⁴ See (Chalmers, 2010).

⁵ On one common interpretation of Kant (the “one world” or “two aspects” interpretation).

collection contribute to a complete picture of a physical world. A collection of experiences is suggestive of trees, we might say, just in case it is a *worldlike* totality of experiences that includes treeish experiences that *cohere* with the other experiences in the totality, where “worldlikeness” and “coherence” are understood as follows.

When you walk through a house, the experiences you have are what we might call “houselike”: they fit together, phenomenologically, in a way that’s analogous to how the frames in a video walk-through of the house would fit together. This is a quality that’s missing from a totality of experiences comprising those you’ve had in the various kitchens you’ve occupied. The most you could achieve by attempting to fit those together in a houselike way would be something analogous to a video montage of various houses’ kitchens.

Just as we can distinguish houselike collections of experiences from non-houselike collections, we can distinguish worldlike collections from non-worldlike collections. A *worldlike* collection comprises experiences of the sort that would characterize the conscious mental lives of beings who collectively explored the whole of some maximal region of space or spacetime. The details needn’t detain us here. The important point is that we do judge ourselves to have explored more or less of actual time and space, based on our experiences; the phenomenology that makes for worldlikeness is the kind that informs such judgements.

Worldlikeness is a property that a totality of experiences can have. Phenomenal coherence is a way that one experience in a totality of experiences can relate to the other experiences in the totality.

An experience *coheres* with a given totality of experiences, just in case it relates to the experiences in that totality as the experiences you’re having now relate to the rest of the experiences you’ve had, rather than as the experiences you have when dreaming or hallucinating relate to the rest of your experiences. (For brevity’s sake, we can also speak of a totality of

experiences containing a “coherent experience,” meaning an experience that coheres with the other experiences in the totality.)⁶

Another important concept for idealism is that of a potential for experience or *phenomenal potential*, as I’ll call it.⁷

It’s uncontroversial that there are phenomenal potentials. Their existence is implicit in our talk of “observable,” “perceptible,” “audible,” “visible,” and “tangible” things. However, the exact nature of phenomenal potentials, like the nature of potentials in general, is controversial. Are potentials (or some of them) ontologically primitive? If not, to what do they reduce? One popular idea is that the existence of a potential reduces to the truth of a suitable subjunctive conditional. Alternatively, we might construe potentials as unconditional probabilities that exceed a certain threshold, or try to analyze them in terms of nomological necessity.⁸

Not all accounts of potential sit equally well with all idealist theories, and this isn’t the place to investigate the choice-points that an idealist faces in this regard. There are, however, three propositions about phenomenal potential that every idealist must accept: (1) all physical things come with associated phenomenal potential (realized or unrealized), (2) facts about phenomenal potential are metaphysically prior to physical facts, and (3) some phenomenal potentials can exist in the absence of anything non-mental (or at least, in the absence of any contingent non-mental entity).

Define the *sensational facts* as the facts about what phenomenal potentials exist in our world. The first tenet of Kantian metaphysics is that any metaphysically possible world that duplicates

⁶ Here and throughout, “experience” is a catch-all for any purely phenomenological event, process, or state of affairs. So, a totality of experiences needn’t be just a set of individual sensations; for example, it might include entire streams of consciousness.

⁷ We could also speak of “phenomenal powers,” “phenomenal tendencies,” or “phenomenal dispositions” in this connection, although disposition-talk strongly connotes something that is disposed, which is inconvenient when it comes to characterizing phenomenalist versions of idealism (see below).

⁸ (Mumford, 1998) and (Molnar, 2003) discuss many of the options in this area.

our world in terms of which sensational facts hold in it is a world in which there hold all of the physical facts that hold in our world (the actual world). Call this tenet of Kantian metaphysics “sensational supervenience.”⁹

Sensational Supervenience: the physical facts supervene on the sensational facts.

Define the *phenomenal field* of a possible world, *W*, as the total phenomenology that would exist in *W*, if all the phenomenal potential that exists in *W* were realized. Then we can put the basic idea behind sensational supervenience like this: our world’s phenomenal field is a worldlike totality of experiences that contains, for every physical feature of our world, a coherent experience as of that feature, and this circumstance is metaphysically sufficient for the existence of that feature.

The second tenet of Kantian metaphysics is that the phenomenal potentials that exist in our world have some categorical basis, meaning that for every phenomenal potential, there is some irreducibly non-modal entity that explains why that potential exists. (By an irreducibly non-modal entity, I mean an entity whose existence doesn’t reduce to the existence of one or more potentials, dispositions, powers, or possibilities.)

The idea behind the second tenet of Kantian metaphysics is that you can’t just have free-floating experience-causing powers: the powers must be powers that something *has*, and this something can’t just be more powers.

Categorical Hypothesis: phenomenal potentials have some categorical basis.

⁹ “Sensational” is a term of art. The sensational facts aren’t necessarily limited to facts about potentials for sensory experience; they might include facts about potentials for other forms of phenomenology, such as cognitive or affective phenomenology. But it’s sensory phenomenology that’s most relevant to a Kantian construction of the physical world.

Kant calls the categorical bases of experience “noumena.” The third tenet of Kantian metaphysics is that the only thing we can know about the noumena is that they exist, and give us various experiences. As Kant puts it, we can know nothing whatsoever about the “inner nature” of the noumena. This is what Rae Langton calls “Kantian humility.”¹⁰

Kantian Humility: the only thing we can know about the categorical basis of phenomenal potential is that such a basis exists and accounts for whatever experiences actually occur.

The final tenet of Kantian metaphysics concerns the nature of conscious experience. Kant doesn't have much to say about this, but it's pretty clear that he doesn't think of consciousness as a physical phenomenon. Anyway, Kantian metaphysics is most interesting when considered as an alternative to physicalism, so it makes sense to include among its tenets one that explicitly excludes physicalism:

Anti-physicalism: the mental doesn't supervene on the physical.¹¹

Kantian metaphysics is the conjunction of these four tenets:

Kantian Metaphysics = Sensational Supervenience + Categorical Hypothesis +
Kantian Humility + Anti-physicalism.

Kant calls his theory “transcendental idealism,” but that's misleading. Idealism is best understood as one of four theories we can classify by how they answer two questions about the relationship between the mental and the physical:¹²

¹⁰ See (Kant, 1781/1998, A277/B333) and (Langton, 1998, 41-43).

¹¹ By calling this tenet “anti-physicalism,” I don't mean to imply that physicalism is equivalent to the claim that the mental supervenes on the physical. The falsity of psychophysical supervenience is sufficient for the falsity of physicalism, but the truth of psychophysical supervenience isn't sufficient for the truth of physicalism.

¹² Fans of grounding can replace the two questions with, respectively, “Does the mental ground the physical?” and “Does the physical ground the mental?”

	Does the physical supervene on the mental?	Does the mental supervene on the physical?
Russellian Monism	Yes	Yes
Idealism	Yes	No
Physicalism	No	Yes
Dualism	No	No

We can't classify the Kantian position according to this scheme, because it's neutral on the question whether the physical supervenes on the mental. All that the Kantian can say is that the physical supervenes on the noumenal, and that this *might* imply that the physical supervenes on the mental, but then again might not: it depends on whether the noumena are mental, which, by Kantian Humility, we can never know.

3 Traditional Idealism

From the standpoint of the philosophy of consciousness, the most we can say about Kantian metaphysics is that it's incompatible with physicalism (by definition), and compatible with an idealist form of monism, but also compatible with a non-monistic world-view. Maybe this is the best we can do (Kant thought so), but one might try to do better, by casting something mental in the role of Kant's noumena. The result would be a frankly monist theory in which mental facts formed the subvenient base of physical reality.

Traditional idealism is just such a theory. It's what you get when you start with Kantian metaphysics, drop Kantian Humility, and replace the Categorical Hypothesis with the stronger *Mental Hypothesis*.

Mental Hypothesis: phenomenal potentials have a mental categorical basis.

Traditional idealism is the conjunction of Sensational Supervenience, the Mental Hypothesis, and Anti-physicalism:

Traditional Idealism = Sensational Supervenience + Mental Hypothesis + Anti-physicalism.

The main proponents of traditional idealism are Leibniz and Berkeley.

Where Kant thinks of physical reality as a potential for certain things (the noumena) to *cause* experiences in suitable patterns, Leibniz thinks of physical reality as a potential for certain things (the monads) to *have* experiences in suitable patterns. For Kant, phenomenal potential has its basis in the unknowable categorical nature of the noumena; for Leibniz, it has its basis in the input-output architecture of the monads.

Leibniz describes the monads as “very exact immaterial automata.” Basically, a monad is a phenomenological Turing machine that takes phenomenal states of the monad as inputs, and returns further phenomenal states of the monad as outputs. The input-output routines that the monads run have neither beginning nor end, so that associated with each monad is an infinitely long stream of consciousness. Furthermore, the monads are isolated, in that no monad affects or is affected by anything else.¹³

Despite their mutual isolation, the mental lives of the monads “harmonize,” in the following sense. You can arrange the monads’ streams of consciousness as rows in a table (one row per monad), in such a way that each column of the table contains a worldlike totality of coherent experiences as of physical things, and successive columns are phenomenal representations of successive stages of a world evolving according to the laws of physics.¹⁴

¹³ See §§1-25 of (Leibniz, 1714/1989a) and §§10-12 of (Leibniz, 1698/1998).

¹⁴ See (Leibniz, 1712/1989), (Leibniz, 1712/2007, 249, 257), and (Leibniz, 1714/1989a, 220). This interpretation of Leibniz’s doctrine of the harmony of the monads takes seriously Leibniz’s view that facts about time, like facts about space and physical objects, reduce to facts about monadic experience (see (Leibniz, 1714/1989b, 307), (Leibniz, 1712/1989, 199), and (Leibniz, 1703/1989, 178)). Since the temporal facts reduce to facts about monadic harmony, we can’t characterize the harmony in terms of a synchrony of monadic experience.

Leibniz was working on the assumption that spacetime was Newtonian, but we could update his theory to accommodate a relativistic account of spacetime. We could say that instead of there being just one collation of monadic experience that yields a picture of a world evolving according to the laws of physics, there are many such collations (one for each foliation of spacetime), where the columns of each collation (each table) correspond to Cauchy surfaces rather than classical time-slices.

Leibniz's idealism is actualistic, in the sense that it takes our world's phenomenal field to consist entirely of *realized* potentials for experience, so that for each physical state of affairs that holds in our world, there are actual experiences as of that state of affairs that cohere with the remainder of monadic experience. In Leibniz's view, no potential for monadic experience goes unrealized.

Leibniz takes this view because he believes that our world is the best possible world, and thinks that the best possible world has to be the one that's richest in fundamental content (and so, in Leibniz's view, richest in monadic experience). But if we set aside these theologically-motivated views, we can consider a potentialist alternative to Leibniz's theory, in which the physical facts of our world supervene on facts about the monads' experience-having potential, there being no assumption now that all (or any) of that potential gets realized.¹⁵

The best-known version of traditional idealism is Berkeley's, according to which all phenomenal potential has its basis in the mind of God.

Like Leibniz, Berkeley is an actualist, in the sense explained above, but Berkeley's commitment to actualism runs deeper than Leibniz's, since Berkeley identifies physical objects

¹⁵ For Leibniz's actualism, see §§57-58 of (Leibniz, 1714/1989a). Leibniz also offers an explanation for the harmony of the monads, proposing that God created them so as to have experiences that harmonize in the way described above. However, one could dispense with this explanation, and just posit the harmony of the monads as a fundamental regularity of the world, analogous to the tendency for certain quantum events occurring in causal isolation from one another to occur in correlated patterns, as in the famous Bell test experiments.

with combinations of experiences. If physical objects are combinations of experiences, you can't very well have the objects without the experiences. Berkeley takes this to be a compelling argument for the existence of God, as a repository for all the physical things that non-divine minds fail to perceive. Most people take it to be a *reductio* of Berkeley.¹⁶

One could modify Berkeley's idealism to escape the commitment to actualism. Berkeley hints at such a modification when he suggests that the existence of an unperceived thing might be a matter of what sorts of experiences *would* occur, if God saw fit to cause certain other experiences. Howard Robinson and John Foster develop potentialist versions of Berkeleyan idealism along these lines.¹⁷

Unlike Kantian metaphysics, traditional idealism is plainly monistic as regards mind and body. But it achieves monism only by replacing Kant's Categorical Hypothesis with the stronger Mental Hypothesis. The question is: why should we accept the Mental Hypothesis?

Each of us knows from his own case that some experiences, and some potentials for experience, have a mental basis, namely a basis in his own mind. It follows that a world in which some phenomenal potentials have their bases in something other than minds is more complicated, ontologically, than a world in which all phenomenal potentials have their bases in minds. Since we should always prefer the simpler of two competing hypotheses, other things being equal, we should prefer the Mental Hypothesis over any alternative hypothesis about the categorical basis of experience.¹⁸

¹⁶ For Berkeley's view that physical objects comprise experiences ("ideas," in his vernacular), see §1 of (Berkeley, 1710/1901). Berkeley denies the existence of matter, but by "matter" he apparently means physical phenomena that don't reduce to anything experiential. Berkeley is best read as a reductionist about the physical, not an eliminativist: see the first ten sections of (Berkeley, 1710/1901).

¹⁷ See §3 of (Berkeley, 1710/1901), (Robinson, 1994, 213-38), and (Foster, 2008, 107-22, 199-245).

¹⁸ See §§26-29 of (Berkeley, 1710/1901). This argument assumes that minds are categorical features of the world. That's debatable—Mill, for example, disagrees (Mill, 1865/1989, 240-49)—but let's set this aside.

The problem with this argument is that it's far from clear that other things *are* equal. The idea that our world consists fundamentally of minds does have a certain simplicity to it, but the simplicity comes at a cost: all those minds.

How high is this cost? Once you've worked your way inside the traditional idealist world-view, it might not look like a cost at all. From the standpoint of the unconverted, however, the benefits of monism aren't sufficient to justify the unexpected introduction of so many minds (or, such a special Mind). Traditional idealists have always touted their theory as the metaphysics of common sense. But common sense is that physical things existed long before there were any minds, and would have existed even if there had never been any minds. Maybe Berkeley would say that this is one point on which common sense must yield to metaphysical insight.

It was precisely because traditional idealism relied so heavily on speculation about the world's mental contents that Kant rejected it in favor of his more cautious position. People sympathetic to traditional idealism might see Kant's caution as a sign of cowardice. Others are more likely to see it as a sober refusal to take a hit from the bong that Berkeley and Leibniz were passing around.

Wherever we come down on this, it's important to recognize that something was lost in the historical shift from theistic idealism to Kantian noumenalism. The shift left Kantians without any satisfying explanation for why there exist the particular potentials for experience that do, in fact, exist. A Kantian can say that these particular potentials exist because the noumena broadcast the particular signals they do, but as long as he adheres to Kantian Humility, this is a purely nominal explanation. It doesn't rationalize the sensational facts of our world in terms of some deeper unifying principle (like a divine plan, carried out directly or via pre-programmed monads). It's like saying that the tide rises and falls for some reason, adding that there's no possibility of learning anything about that reason, besides that it accounts for the tide's rise and fall.

Physicalists and dualists take various physical entities as the categorical basis of experience, and explain why experiences tend to occur as they do by reference to physical features of those entities. Traditional idealists take various minds as the categorical basis of experience, and explain why experiences tend to occur as they do by reference to mental features of those minds. But a Kantian can't say anything illuminating about the supposed categorical basis of experience. All he can say about the noumena is that they exist—the ghosts of departed deities.

4 Phenomenalism

One reason to posit categorical bases for phenomenal potentials is to explain why there exist the particular potentials that do; however, as we've just seen, this isn't a reason that's available to Kantian metaphysicians. If Kantians have a reason to posit a categorical basis for the phenomenal potentials of our world, it can only be that it's in the very nature of potentials to have categorical bases.

The idea that potentials, tendencies, or dispositions—"modality," for a catch-all—must have categorical bases was an important doctrine of late 20th century metaphysics. David Armstrong was probably the doctrine's staunchest advocate, insisting throughout his long career that if a potential exists, it must be due to the existence of something that isn't a potential. Call this the "Armstrong Doctrine."¹⁹

If the Armstrong Doctrine is true, we have no choice but to accept the Categorical Hypothesis. Whether it *is* true is a matter of ongoing debate; however, there are reasons to doubt it.

The possibility of potentials that have no categorical basis—"base-free" potentials, as I'll call them—would appear to be established by the possibility that our own world is fundamentally

¹⁹ See (Armstrong, 1961, 56-58) and (Armstrong, 1993, 187); also (Lewis, 1992, 218-19) and (Lewis, 1998).

chancy. As far as we know, it's a physically fundamental fact that there's about a 50% chance that the quantum tunneling involved in the decay of a radon atom occurs within four days of the genesis of the atom. Presumably, about 50% of the radon atoms that come into existence in our world decay within four days. But we can imagine a world categorically indistinguishable from ours, in which there's only about a 10% chance that a radon atom decays within four days of its genesis. We need only imagine that due to a colossal statistical fluke, about 50% of the radon atoms in this other possible world decay within four days, despite there being only about a 10% chance that any given radon atom decays within four days. In this other world, there exist potentials for radioactive decay that do not exist in our world, despite the worlds' being categorically indistinguishable.²⁰

One philosopher who definitely would have rejected the Armstrong Doctrine is J.S. Mill. Unlike Leibniz, Kant, and Armstrong, Mill sees no reason to think that our world isn't potentials all the way down. This leads Mill to reject Kant's Categorical Hypothesis (and with it the stronger Mental Hypothesis of traditional idealism), and to eschew Kantian Humility in favor of what we might as well call Millian Humility.

Millian Humility: for all we know, phenomenal potentials have no categorical basis.

In effect, Mill out-Kants Kant: whereas Kant assumes the existence of a categorical basis for experience and merely suspends judgement as to its more particular categorical nature, Mill suspends judgement on the question of whether a categorical basis for experience exists at all.

²⁰ Stephen Mumford argues for the reality of "ungrounded" (i.e., base-free) dispositions along these lines; see (Mumford, 2006). Jennifer McKittrick argues for the metaphysical possibility of bare dispositions in (McKittrick, 2003); her arguments translate naturally into arguments for the metaphysical possibility of base-free potentials.

Phenomenalism is what you get when you start with Kantian metaphysics, drop the Categorical Hypothesis, and replace Kantian Humility with Millian Humility. It's the conjunction of Sensational Supervenience, Millian Humility, and Anti-physicalism:

Phenomenalism = Sensational Supervenience + Millian Humility + Anti-physicalism.

Just as you can think of Berkeleyan idealism as the result of replacing Kant's noumena with Berkeley's God, and Leibnizian idealism as the result of replacing Kant's noumena with Leibniz's monads, you can think of Millian phenomenalism as the result of replacing Kant's noumena with Mill's "permanent possibilities of sensation"—potentials for experience that might have categorical bases, but might not, and can serve as the subvenient base of physical reality either way.²¹

Unlike Kantian metaphysics, which is compatible with idealism but also with its denial, phenomenalism is best classified as an idealist theory. This is because phenomenologists, unlike Kantians, don't put anything into their fundamental ontology that they don't classify as mental. A phenomenologist allows that there might be (unbeknownst to us) something non-mental that somehow explains phenomenal potentials, but, if so, its relationship to the potentials is analogous the relationship of physical brains to conscious experiences in dualist theories—a relationship that's consistent with classifying the experiences as purely mental.

²¹ For Mill's phenomenalism, see Chapter XI of (Mill, 1865/1989), and the Appendix to Chapters XI and XII. Other sympathetic discussions of phenomenalism include (Price, 1932), (Lewis, 1946, 203-53), (Ayer, 1946-1947), (Fumerton, 1985, 131-73), and (Pelczar, 2015). If we replace Millian Humility in the definition of phenomenalism offered above with the claim that the sensational facts have no categorical basis, we get a theory we might call "ambitious phenomenalism." The problem with ambitious phenomenalism is that it's hard to see what justifies the claim that the sensational facts have no categorical basis. If Mill is right, considerations of theoretical simplicity compel us not to posit such a basis, but that's different from positing its non-existence.

It's true that phenomenalism is a kind of idealism only if the existence of a potential for experience is a mental state of affairs. But since we classify potentials for physical events (like radioactive decay) as physical, it seems reasonable to classify potentials for experience as mental.

It's also true that in the phenomenalist view, facts about potential experience underdetermine facts about actual experience (and vice versa). But in the traditional idealist view, facts about minds underdetermine facts about actual experience (and vice versa), and this doesn't deter us from counting traditional idealism as a kind of monism.

A commitment to phenomenalism doesn't carry with it a commitment to suspending judgement on whether phenomenal potentials have any explanation. It just carries a commitment to suspending judgement on whether they have any *categorical* explanation. It's consistent with phenomenalism for one sensational fact (or collection of sensational facts) to explain another sensational fact.

Actually, a phenomenalist has to allow for this kind of explanation, in order to avoid the absurdity of holding that physical facts are universally inexplicable.

The existence of a delta at the mouth of the Mississippi River isn't a miracle. It's the result of thousands of years of silt- and sand-deposits occurring where the river slows as it enters the Gulf of Mexico. Like anyone else, a phenomenalist recognizes that the delta is a natural consequence of these hydrological processes. It's just that a phenomenalist sees both the delta and the processes that created it as *metaphysical* consequences of the existence of various potentials for experience.

The motions of water and sediment are a metaphysical consequence of certain sensational facts, the delta is a metaphysical consequence of certain other sensational facts, and the latter sensational facts are a natural, non-metaphysical consequence of the former sensational facts. For a phenomenalist, to say that the delta is a "natural consequence" of the hydrological

circumstances (the water flow, the sediment, etc.) is to say that it's a natural law, or a consequence of natural laws, that if the sensational facts that entail those hydrological circumstances obtain, so do sensational facts that entail the existence of a delta.²²

A phenomenalist holds that many (perhaps all) sensational facts have non-reductive explanations *in terms of other sensational facts*. This is compatible with Millian Humility, which only requires us to suspend judgement on whether any sensational fact is based in (or explained by) something that isn't itself a sensational fact.

Phenomenalism avoids the ontological extravagance of traditional idealism. In Mill's view, the existence of a physical object doesn't require the existence of a mind capable of having or causing experience, any more than the existence of a Kantian noumenon requires the existence of a mind capable of receiving its signals. Nor, in Mill's view, does the existence of a physical object require any actual experience. According to Mill, the natural world is just a big potential for experience, some of which happens to be realized, but most of which is not (at least, as far as we know).

A theory can be more or less simple along either of two dimensions. One dimension of simplicity is ontological: the less fundamental stuff a theory gets by with (or, the fewer fundamental *kinds* of stuff), the better, all else being equal. The other dimension of simplicity is architectural: a theory that gets by with less-complicated fundamental laws is better, other things being equal, than a theory that requires more complicated fundamental laws.

Ontologically, phenomenism is very simple. How about architecturally?

²² As Mill puts it, "Whether we are asleep or awake the fire goes out, and puts an end to one particular possibility of warmth and light. Whether we are present or absent the corn ripens, and brings a new possibility of food. Hence we speedily learn to think of Nature as made up solely of these groups of possibilities, and the active force of Nature as manifested in the modification of some of these by others." (Mill, 1865/1989, 230) See also (Ayer, 1940, 229-31) and (Ayer, 1946-1947, 146-50).

For fundamental laws, physicalism gets by with the laws of physics alone. Taking this as our baseline, we see that dualism is architecturally more complex than physicalism, but not dramatically so: it just supplements the laws of physics with some psychophysical bridge laws, to account for observed correlations between experiences and their neural correlates. Traditional idealism is on an architectural par with physicalism; it's just that instead of taking the laws of physics as fundamental, it takes as fundamental the principles that govern the underlying mental reality on which the laws of physics supervene.

Like the traditional idealist, the phenomenalist does not regard the laws of physics as metaphysically fundamental. But unlike the traditional idealist, the phenomenalist can't base the laws of physics on a deeper level of minds governed by laws that are metaphysically fundamental. What *can* the phenomenalist offer by way of fundamental laws?

We're told that it's a law of physics that for every action, there's an equal and opposite reaction. If that's true, a phenomenalist will say that it's a law of experience that if the phenomenal field contains a coherent experience of an action, it contains a coherent experience of an equal and opposite reaction. Likewise for all other laws. If the Einstein Field Equations express a physical law, then it's a law of experience that the phenomenal field contains coherent experiences of whatever physical phenomena inspire the physicists who contemplate them to accept the Einstein Field Equations.²³

Where the physicalist has laws of physics, and the traditional idealist laws of minds (or Mind), the phenomenalist has laws of experience.

²³The reference to physicists' contemplations here is only to gesture towards the actual content of the relevant law of experience; the law itself would make no reference to physicists or their contemplations.

We can state the laws of physics in terms of a few dozen (at most a few hundred) properties, relations, and natural kinds—physically fundamental fields, forces, particles, etc. Is a similar economy of terms possible in the expression of a phenomenalistic law of experience?

It depends on how rich the phenomenal field must be, in order to serve as the subvenient base of the physical regularities that correspond to physical laws. If it's enough for the field to contain experiences characterized by a smallish range of qualia—perhaps the qualia by virtue of instantiating which our experiences present things as having various geometric and temporal features—then phenomenalism may be on a par with physicalism or traditional idealism, architecturally. If the phenomenal field must be significantly richer than that to support the regularities that physical laws describe, phenomenalism may be at an architectural disadvantage to alternative metaphysical schemes.

5 Challenges to sensational supervenience

We've saved the biggest challenges to idealism for last. These are objections to sensational supervenience, the central tenet of all idealist metaphysics (and of Kantian metaphysics too). We can state the objections as conceivability arguments against sensational supervenience, of which we'll consider three.²⁴

Consider first the *Matrix Argument*:

We can conceive of a world in which there holds every sensational fact that holds in our world, but in which those facts hold only because of the operations of a supercomputer connected to some envatted brains. Furthermore, we can conceive

²⁴ Conceivability arguments are controversial, but a main motive for taking idealism seriously is dissatisfaction with physicalism, and a main source of dissatisfaction with physicalism is its vulnerability to conceivability arguments against psychophysical supervenience. So, an idealist has to take conceivability arguments against his own theory seriously.

of this world—call it Matrix World—as being physically very different from ours; e.g., as containing no trees. This gives us a compelling reason to believe that the sensational facts about our world (the actual world) do not metaphysically entail the physical facts about our world.

An idealist can respond to this argument by granting the whole thing.

Sensational supervenience says that any metaphysically possible world that is indistinguishable from ours with respect to the sensational facts that hold in it is a world that has all the physical features that our world has. A counterexample to sensational supervenience would be a metaphysically possible world characterized by all *and only* the sensational facts that characterize our world (the actual world), but that lacked some of our world's physical features.

Matrix World is not such a world: in Matrix World, there are many sensational facts that don't hold in our world. For example, in Matrix World, the phenomenal field includes coherent experiences as of a powerful computer attached to some brains; this is a sensational fact that presumably doesn't hold in our world (or if it does, then Matrix World might duplicate ours physically after all, in which case the argument collapses).

What if we suppose that the supercomputer, vats, and related paraphernalia of Matrix World are for some reason imperceptible? Suppose we stipulate that in Matrix World, nothing can perceive the supercomputer and so forth “from the outside” (i.e., other than in the ways that the envatted brains perceive the computer, if they can be said to perceive it). This version of Matrix World still fails to contain trees, but unlike the earlier version, it contains no more phenomenal potential than our world, and therefore escapes the idealist come-back offered above.

The idealist response to this is that we can't conceive of such a world.

We can conceive of a world that contains vats, supercomputers, etc. despite containing no conscious experience. (Berkeley thought otherwise, for reasons that are pretty clearly fallacious.) The question here, though, is whether we can conceive of a world that contains vats, supercomputers, etc. despite containing no *potential* for experience.²⁵

It's hard to do justice to this question without saying more about the nature of phenomenal potential than there's room to say here. Still, there's reason to be skeptical about the suggestion that we can conceive of a world that contains vats and so forth but no potential for relevant experiences (as of vats and so forth).

If we want to conceive of a world that contains vats without any corresponding potential for experience, how should we go about it? We could imagine some kind of cloaking device that causes anyone who wanders into the vats' vicinity to hallucinate an absence of vats. But then the cloaking device would itself have to be imperceptible, in order for its presence not to entail a difference between Matrix World and our world, at the level of sensational facts.²⁶

A better way to try to conceive of an imperceptible vat without simply shifting the focus of discussion (to an imperceptible cloaking device, or whatever) is by trying to conceive of it in purely structural terms. If, as proponents of ontic structural realism contend, the existence of any physical entity reduces to the satisfaction of some purely structural description—a description, like “ $\exists\phi\exists x(\phi x)$,” that employs only logical and mathematical terms—then one could, in theory, conceive of a world containing vats just by conceiving of a world that satisfies certain purely structural descriptions. Assuming, plausibly, that we can conceive of a purely structural

²⁵ Even if vats and other middle-sized dry goods can't exist in the absence of a corresponding potential for experience, there might be other physical entities (like subatomic particles) that can; we consider the challenge that such entities (or alleged entities) pose to idealism below.

²⁶ What about a self-cloaking device that ensures that the only experiences that occur are ones that contribute to a totality suggestive of a device-free world physically indistinguishable from our own? There is room for such a device in an idealist world-view. Kantians call it a noumenon.

description being satisfied in the absence of any potential for experience, it would follow that we can conceive of a counterexample to sensational supervenience by conceiving of the vats and so forth in Matrix World in purely structural terms, and stipulating that there is no corresponding potential for experience.²⁷

This isn't the place to debate the merits of structuralist metaphysics. The important point is that there is a real threat to idealism here, and that meeting it requires a close engagement with the structuralist world-view. At the end of the day, idealists might have more to fear from structuralism than from physicalism, dualism, or Russellian monism.²⁸

A different modification of the Matrix Argument replaces the envatted brains with disembodied minds, and the supercomputer with interactions among those minds. Thus we have the *Ghost Argument*:

We can conceive of a world consisting of a multitude of disembodied minds; call it Ghost World. The minds are capable of interactions that result in their having various experiences, and these interactions, or potential interactions, are governed by laws that determine the patterns in which the resulting experiences occur. The laws and minds are such that all and only the phenomenal potentials that exist in our world (the actual world) exist in Ghost World. Since Ghost World contains nothing physical, despite duplicating our world at the level of sensational facts, our ability to conceive of Ghost World gives us a compelling reason to believe that the physical facts don't supervene on the sensational facts.

²⁷ For structuralism about the physical, see (Russell, 1927), (Ladyman *et al.*, 2007), and (Tegmark, 2014). The general idea goes back at least to (Boscovich, 1763/1922).

²⁸ That said, proponents of sensational supervenience have reason for optimism *vis à vis* the structuralist threat, since structuralism about the physical faces a serious objection due to Max Newman. In effect, Newman shows that structuralism, at least in its purest form, implies that it's impossible for two worlds that contain the same number of things to differ from one another physically; see (Newman, 1928).

In response, an idealist can deny that there's any physical difference between Ghost World and the actual world.

From an idealist standpoint, the difference (assuming there is one) between our world and Ghost World isn't that our world but not Ghost World contains physical things. The difference is in what ultimately explains why there exist the physical things that do—these being the same physical things (or indistinguishable physical things) in the case of each world. In Ghost World, the existence of physical things has its ultimate explanation in certain lawlike interactions among various minds; in a Leibnizian world, it has its ultimate explanation in a Divine Plan; in a Kantian world, it has its ultimate explanation in the causal powers of the noumena; in our world, it has its ultimate explanation in—well, who knows? Maybe nothing.

This response might not be available to all idealists. Berkeley contends that most physical things in our world consist of divine experiences. If, as many hold, it's essential to a physical object that it consist of whatever sort of stuff it actually consists of, then Ghost World doesn't contain all the physical things that exist in our world, since Ghost World contains no divine experiences. But if we equate a physical object with a combination of phenomenal potentials, and individuate such potentials by the experiences for which they are potentials, rather than by the potentials' categorical bases (if any), we can say that Ghost World does contain the same physical things as our world, since it contains the same potentials, even if the potentials have a categorical basis in Ghost World that they don't have in ours.

A final argument against sensational supervenience targets a perennial source of dissatisfaction with idealist metaphysics, which is its treatment of unobservable phenomena. Call it the *Argument from Unobservables*:

We can conceive of a world, WYSIWYG (“what-you-see-is-what-you-get”) World, that has all the same observable features as our world, but no unobservable features. Since WYSIWYG World has the same observable features as our world, it includes the same potentials for experience as our world. But since WYSIWYG World lacks the unobservable things that exist in our world (subatomic particles, etc.), it differs from our world physically. So WYSIWYG World is a counterexample to sensational supervenience.

One possible idealist response is to allow that our world has whatever unobservable features science tells us it has, but deny that a world could duplicate ours sensationally without having all those features. The idea here is that the distinction between observable and unobservable phenomena isn’t a distinction between phenomena for which there are corresponding potentials for experience and phenomena for which there are not, but rather a distinction between two kinds of potential for experience.

We can put this in terms of the phenomenal field. An observable thing, like the Statue of Liberty, is a conspicuous pattern in the phenomenal field. It’s a pattern that consists of experiences that all resemble one another in obvious ways—they’re all experiences as of a statue with a certain shape, size, color, etc. An unobservable thing, like an electron or gravitational wave, is an inconspicuous pattern in the phenomenal field. It’s a pattern consisting of experiences that do not resemble one another in any obvious way, but exhibit some order, symmetry, or regularity that emerges when we subject the experiences to an appropriate mathematical description. In this view, unobservable phenomena are, so to speak, hidden patterns in experience.²⁹

²⁹ Theoretical physicists obviously don’t work directly from the experiences of observational physicists: they work primarily from the observationalists’ records of their observations. But those records are a reflection of the

An alternative response to the argument from unobservables is to deny that there are any unobservable things (or at least, any that a hidden patterns account can't handle).

One way to do this would be to hold that although many things are unobservable to us humans, no physical entity is absolutely unobservable, since (one might argue) there's always the possibility of a Laplacean Demon who observes subatomic particles, the interiors of black holes, etc.

But what if our world contains physical entities that aren't just unobservable to this or that sentient being, but unobservable in principle? (Perhaps one could argue that subatomic particles and the interiors of black holes are such things.) And what if it's impossible to construe these in-principle unobservable things as hidden patterns in experience?

An idealist has to deny that such things exist. Is this a problem for idealism?

Thinking of the world as containing various unobservable things obeying various rules definitely helps us make sense of what we observe. But it doesn't follow that the world actually contains unobservable things. It might be that the alleged unobservables are really just accounting devices, like the international dollar, or, if you like, fictional characters in the scientific narrative.

The suggestion is far from new: it's scientific antirealism.

Scientific antirealism is controversial, but an idealist might not have to buy into it lock, stock, and barrel in order to overcome the argument from unobservables. He can adopt a wait-and-see policy of suspending judgement on whether unobservable things exist *until and unless we can make good idealistic sense of them*, either by forming a clear conception of what it would be like to observe the things, or by finding a way to construe the things as hidden patterns in experience.

observationalists' experiences, and by discovering a pattern in the records, the theoretician implicitly discovers a pattern in the corresponding experiences.

Before there were microscopes, people couldn't observe microbes. Still, a visionary Renaissance doctor might have proposed a germ theory of disease, positing germs to explain the transmission and progression of various illnesses.

We can imagine a debate arising over the reality of these so-called "germs," with some people inclined to think that they're a genuine biological phenomenon, and others inclined to think of them as convenient fictions.

The germ-antirealists will be forced to admit they're wrong, when people start actually observing germs through microscopes. But if they adopt the wait-and-see policy described above, there's no harm done: what they used to regard as convenient fictions, they can now regard as realities supervening on sensational facts involving (among others) the kinds of experiences people have when looking through microscopes.

Similarly, if there comes a time when we can make sense of perceiving a quark (assuming we can't make sense of this already), an idealist can give an account of quarks in terms of the phenomenology of such perceptions. Until then, or until he finds a way to construe quarks as hidden patterns in the phenomenal field, an idealist can treat them as convenient fictions.

Is the wait-and-see policy a reasonable one? That's debatable, but it may have to be, if idealism is to survive the argument from unobservables.

6 Conclusion

It's possible that idealism will always be one of those theories that looks compelling from the inside, but implausible from the outside. If so, it's in reputable company: physicalism, arguably, is also such a theory. For those who prefer a metaphysics that looks equally good from all angles, at the cost of not looking particularly stunning from any angle, there's always dualism.

Today, of course, the philosophy of consciousness is almost entirely a conversation between physicalists and dualists. What would it take to get idealism back into the game?

For one thing, it would help to dissociate idealism from some questionable arguments that people have made for it.

One such argument goes like this: we know that there is an external world; idealism is the only metaphysics that is compatible with our having such knowledge; therefore, idealism is true.³⁰

The problem here is the second premise. Like any defensible theory, idealism has to recognize a distinction between veridical and non-veridical experience. A skeptical hypothesis for idealism is just one in which all of our experiences fall into the latter category—i.e., in which the phenomenal field contains no experiences that cohere with the other experiences in the field. There is nothing in idealism to rule this out (or at least, nothing that wouldn't serve to rule out the corresponding possibility in any other metaphysical setting). If physical objects are woven into a veil of ideas, it's only because the veil has an overall pattern that it could fail to have, consistent with all our actual experience.³¹

Another traditional argument for idealism goes like this: the world (or its non-abstract part) consists ultimately of its categorical contents; the only non-abstract things capable of having categorical natures are conscious minds or experiences; therefore, the world (or its non-abstract part) consists ultimately of conscious minds or experiences.³²

³⁰ See §§87-91 of (Berkeley, 1710/1901) and Part IV of Book I of (Hume, 1739/1978).

³¹ There might be a better epistemic argument for idealism, though: see (Smithson, 2017).

³² See (Leibniz, 1686/1998), (Eddington, 1928, 247-72), (Hartshorne, 1946, 413), (Adams, 2007, 40), (Foster, 2008, 42-82), and (Strawson, 2008, 19-51).

This argument is better than the epistemic one, but hardly compelling. Not even all idealists agree with its first premise (recall Mill), and convincing arguments for its second premise are elusive.³³

The best argument for idealism is probably this: some form of idealism is the simplest defensible metaphysics; we should accept the simplest defensible metaphysics; therefore, we should accept some form of idealism.

Making this argument work requires defending idealism from the principal objections to it, and showing that idealism is simpler than any defensible alternative. That's a tall order, but nothing less is likely to put idealism back on the map.³⁴

³³ If there is a good argument for the second premise, it's probably along the lines of (Robinson, 1982, 108-23).

³⁴ Thanks to Bob Beddor, Ben Blumson, Brian Cutter, Uriah Kriegel, Abelard Podgorski, Qu Hsueh Ming, Neil Sinhababu, and Rob Smithson for their comments on earlier drafts of this chapter.

References

Adams, Robert M. 2007. Idealism vindicated. *Pages 35-54 of: Persons: Human and Divine.*

Oxford: Oxford University Press.

Armstrong, David. 1961. *Perception and the Physical World.* London: Routledge & Kegan

Paul.

Armstrong, David. 1993. Reply to Martin. *Pages 186-194 of: John Bacon, Keith Campbell, &*

Reinhardt, Lloyd (eds), *Ontology, Causality, and Mind: Essays in Honour of D.M.*

Armstrong. Cambridge: Cambridge University Press.

Ayer, A.J. 1940. *The Foundations of Empirical Knowledge.* London: Macmillan.

Ayer, A.J. 1946-1947. Phenomenalism. *Proceedings of the Aristotelian Society*, **47**, 163-196.

Berkeley, George. 1710/1901. A Treatise Concerning the Principles of Human Knowledge.

Pages 233–347 of: The Works of George Berkeley, A.C. Fraser, ed., vol. Vol. 1. Oxford:

Clarendon Press.

Boscovich, Roger Joseph. 1763/1922. *A Theory of Natural Philosophy, J.M. Child, trans.*

Chicago & London: Open Court Publishing Company.

Chalmers, David. 2010. The Matrix as metaphysics. *Pages 455-494 of: The Character of*

Consciousness. Oxford: Oxford University Press.

Eddington, A.S. 1928. *The Nature of the Physical World.* New York: Macmillan.

Foster, John. 2008. *A World for Us: The Case for Phenomenalistic Idealism.* Oxford: Oxford

University Press.

Fumerton, Richard A. 1985. *Metaphysical and Epistemological Problems of Perception.*

University of Nebraska Press.

Hartshorne, Charles. 1946. Leibniz's greatest discovery. *Journal of the History of Ideas*, 7(4), 411-421.

Hume, David. 1739/1978. *A Treatise of Human Nature*. Oxford: Clarendon Press.

Kant, Immanuel. 1781/1998. *Critique of Pure Reason*. Cambridge: Cambridge University Press.

Ladyman, James, Ross, Don, & David Spurrett, with John Collier. 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Langton, Rae. 1998. *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford: Clarendon Press.

Leibniz, G.W. 1686/1998. Leibniz to Arnauld, 4/14 July 1686. *Pages 105-114 of: G.W. Leibniz: Philosophical Texts, translated by R. Francks and R.S. Woolhouse*. Oxford: Oxford University Press.

Leibniz, G.W. 1698/1998. A Letter from M. Leibniz to the Editor, Containing an Explanation of the Difficulties Which M. Bayle Found with the New System of the Union of the Soul and Body. *Pages 201-208 of: G.W. Leibniz: Philosophical Texts, translated by R. Francks and R.S. Woolhouse*. Oxford: Oxford University Press.

Leibniz, G.W. 1703/1989. Leibniz to de Volder, 20 June 1703 [excerpts]. *Pages 174-178 of: Ariew, Roger, & Garber, Daniel (eds), Philosophical Essays*. Indianapolis and Cambridge: Hackett Publishing Company.

Leibniz, G.W. 1712/1989. Notes for Leibniz to Des Bosses, 5 February 1712. *Pages 199-200 of: Ariew, Roger, & Garber, Daniel (eds), Philosophical Essays*. Indianapolis and Cambridge: Hackett Publishing Company.

Leibniz, G.W. 1712/2007. Leibniz to Des Bosses, 16 June 1712. *Pages 255-257 of: Look, Brandon C., & Rutherford, Donald (eds), The Leibniz-Des Bosses Correspondence*. New Haven: Yale University Press.

- Leibniz, G.W. 1714/1989a. *Monadology*. *Pages 213–225 of*: Ariew, Roger, & Garber, Daniel (eds), *Philosophical Essays*. Indianapolis and Cambridge: Hackett Publishing Company.
- Leibniz, G.W. 1714/1989b. *Remarks on Berkeley's Principles*. *Page 307 of*: Ariew, Roger, & Garber, Daniel (eds), *Philosophical Essays*. Indianapolis and Cambridge: Hackett Publishing Company.
- Lewis, Clarence Irving. 1946. *An Analysis of Knowledge and Valuation*. La Salle: Open Court.
- Lewis, David. 1992. Critical notice of *A Combinatorial Theory of Possibility*. *Australasian Journal of Philosophy*, **70**(2), 211-224.
- Lewis, David. 1998. The truthmakers. *Times Literary Supplement*, **4948**, 30.
- McKittrick, Jennifer. 2003. The bare metaphysical possibility of bare dispositions. *Philosophy and Phenomenological Research*, **66**(2), 349-369.
- Mill, John Stuart. 1865/1989. *An Examination of Sir William Hamilton's Philosophy, and of the Principal Philosophical Questions Discussed in his Writings*. London: Longman, Green, and Co.
- Molnar, George. 2003. *Powers: A Study in Metaphysics*. Oxford: Oxford University Press.
- Mumford, Stephen. 1998. *Dispositions*. Oxford: Oxford University Press.
- Mumford, Stephen. 2006. The ungrounded argument. *Synthese*, **149**(3), 471-489.
- Newman, M.H.A. 1928. Mr. Russell's "Causal Theory of Perception". *Mind*, **37**(146), 137-148.
- Pelczar, Michael. 2015. *Sensorama: A Phenomenalist Analysis of Spacetime and Its Contents*. Oxford: Oxford University Press.
- Price, H.H. 1932. *Perception*. London: Methuen & Co.
- Robinson, Howard. 1982. *Matter and Sense: A Critique of Contemporary Materialism*. London: Cambridge University Press.
- Robinson, Howard. 1994. *Perception*. London: Routledge.

- Russell, Bertrand. 1927. *The Analysis of Matter*. London: Kegan Paul, Trench, Trubner & Co.
- Smithson, Robert. 2017. A new epistemic argument for idealism. *Pages 17-33 of:* Goldschmidt, Tyron, & Pearce, Kenneth L. (eds), *Idealism: New Essays in Metaphysics*. Oxford: Oxford University Press.
- Strawson, Galen. 2008. *Real Materialism: and Other Essays*. Oxford: Clarendon Press.
- Tegmark, Max. 2014. *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*. New York: Alfred A. Knopf.