Algorithmic AI Consciousness

Abstract

I argue that the thoroughly algorithmic nature of current AI systems (such as LLMs) is no obstacle to their being conscious. To this end, I present a picture on which current AI systems comprise dispositional properties which realize categorical phenomenal properties where the latter, in turn, provide the identity conditions for their dispositional realizers. This mutual ontological dependence, or, symmetrical grounding, at the heart of the proposal yields a novel picture of (AI) consciousness that avoids epiphenomenalism and is more permissive regarding the specific nature/functional organization of conscious systems than has been previously suggested. This, in turn, suggests an epistemology of AI consciousness focused on investigating the high-level behaviours of AI systems rather than their low-level functional organization.

Key Words: Algorithms; Artificial Intelligence; Consciousness; Dispositions; Grounding; Large Language Models

1. AI is Algorithmic

Current AI models are built on algorithms—structured sets of instructions that guide the system from one state to the next by defining how the AI processes a variety of inputs to produce corresponding outputs. There are multiple ways to implement this foundational algorithmic design. Large language models (LLMs), for example, function as advanced pattern recognizers. They generate "intelligent" responses by predicting the most likely next words based on the given input.[1] This process involves several key steps. For example, each word in the input is assigned a numerical representation, which defines its position within a multidimensional space where the distances between points reflect similarities in meaning. Next, the model assigns each word an "attention score"—a precise value determined by a mathematical formula— to assess its importance to the overall context. Additional algorithmic layers further refine the system's understanding (*viz.* its placement of words in "meaning space" and assignment of attention score values) before it ultimately predicts the most probable next token, producing the final output. At every stage, the model operates according to carefully defined rules that dictate how outputs are generated from inputs. In short, AI models are fundamentally and thoroughly algorithmic in nature.

---

[1] At one salient level of description, it is true to say that LLMs are "pattern recognizers" or "next token predictors". But there are plausibly higher-level behavioural dispositions in terms of which these systems can be accurately described too (Grzankowski, Downes, and Forber Forthcoming).

In advanced AI systems, high-level outputs for a given input can be unpredictable. Moreover, such outputs (e.g., text generation in large language models (LLMs)) are not the result of direct, output-specific programming. Yet, this unpredictability does not challenge the fundamentally algorithmic nature of these systems. Rather, the unpredictability arises precisely because of the complex interdependencies among the underlying algorithms. From the user's high-level perspective, this complexity manifests as novel and surprising outputs. Crucially, however, this novelty remains explicable within the system's low-level algorithmic architecture. As Patterson and Gibson (2017, 1) observe:

> Fundamentally, machine learning is using algorithms to extract information from raw data and represent it in some type of model. We use this model to infer things about other data we have not yet modelled.

Expanding on this, Coates (2023, 183–84) explains:

> The ultimate outcomes of this sort of system are determined by a generally highly complex process in which models are constructed from data and then used to generate outputs in response to novel inputs. So, the system is not initially programmed to produce specific outputs, and its outputs often cannot be predicted.

Thus, the apparent unpredictability and emergent behaviour of these systems do not signify a departure from algorithmic foundations. On the contrary, they exemplify the power and depth of algorithmic processes—processes that remain, from the lowest computational level to the highest behavioural output, algorithmic through and through.

2. Categoricalism and Dispositionalism

In debates about the metaphysics of properties, a distinction is drawn between *dispositional* properties and *categorical* properties. The dispositional properties of an individual concern what it *would* do in certain circumstances. For example, *fragility* is a dispositional property, a fragile vase is such that it *would* shatter if dropped. In other words, dispositional properties are *modal* since they concern alternative possibilities for the *disposed* objects. Categorical properties, on the other hand, do not primarily concern what an object *would* do, given the appropriate stimulus, they are…*categorical* rather than conditional in nature. Categorical properties don't essentially "point beyond themselves" to some merely possibly manifestation condition. Geometrical properties are plausible examples of categorical properties because their *essences*—what it is to be the property in question—can be fully specified without invoking any modal notions. The property *sphericity*, for example, is the property of having a surface

all points on which are equidistant from a given point (see, e.g., Lowe 2010; Yates 2018).

It is uncontroversial to recognise this dispositional/categorical divide among properties. There is, however, plenty of controversy surrounding the following "-*isms*":

> **Categoricalism:** Dispositional properties require categorical realizers.
> **Dispositionalism:** Dispositional properties do not require categorical realizers.

To get an intuitive sense of what's at stake here, consider the fact that while the property *sphericity* is categorical, it also realizes dispositions such as a disposition to roll down an incline, to cast an elliptical shadow, to make a concave impression in sand, and so on. So, we might say that the categorical property, *sphericity*, realizes these dispositional properties. The debate between Categoricalism and Dispositionalism is thus a debate about whether this is true for *all* properties.

Dispositionalism is particularly attractive when we consider just the low-level physical properties such as *charge, mass*, and *quantum spin*. We specify the nature of *charge* (for example) in terms of what it disposes its bearers to do in certain experimental situations. The dispositionalists then maintain that there is no need to also posit a categorical nature underlying these dispositions, such additional structure would be causally idle and thus unknowable in principle and so is not the sort of thing we should include in ontology. Dispositionalists argue that *charge*, like other low-level physical properties, is *purely* dispositional; in other words, they say that the *essence* of the property *charge* is exhausted by dispositional relations to other properties (see, e.g., Mumford 2004; Bird 2007; Chakravartty 2007). Categoricalists, by contrast, maintain that *charge*, like all properties, is fundamentally and essentially *categorical*. This categorical nature of the property, which may be as "thin" as mere primitive self-identity and distinctness from other properties, then realises the dispositions documented by science, perhaps with the help of *laws of nature* (see, e.g., Armstrong 1983; Lewis 2009).

3. The "Intuitive View"

These issues in the metaphysics of properties are relevant to the metaphysics of *consciousness* because there is a prevalent view in the philosophy of mind according to which phenomenal properties, such as *being in pain, feeling hungry, tasting an ice cream*, are categorical. Phenomenal properties are not purely or primarily dispositional, like *charge* plausibly is, rather, they are primarily specified in terms of how they *feel* to us. The categorical feels of phenomenal properties constitute their *essences*. This is consistent with phenomenal properties nonetheless being systematically associated with behavioural dispositions. Pain, for example, is primarily defined and understood

in terms of how it feels to us, but this categorical *feeling* of pain also determines certain behavioural dispositions such as recoiling when touching a hot surface and anxious sweating at the prospect of an injection.

What's more, the categorical phenomenal nature of pain, to which we have direct epistemic access, arguably renders its determination of certain behavioural dispositions fully intelligible; there's no *mystery* regarding the connection between the pain of feeling a hot surface and the resultant recoiling behaviour.[2] This *Intuitive View* (cf. Coates 2023, 170–71) of the categorical nature of phenomenal properties in relation to behavioural dispositions is in the background of discussions across the spectrum of the metaphysics of mind, from the physicalist "phenomenal concepts" strategy (see, e.g., Stoljar 2005; Papineau 2006; Balog 2012) to panpsychism (e.g., Strawson 2006; Mørch 2018; Goff 2019) and Cartesian interactionist dualism. The phenomenal concepts strategists cite the fact that we primarily conceptualize phenomenal properties in terms of their categorical "feels" in what they claim to be a phsicalistically acceptable explanation of the intuition of an explanatory gap between the physical and the mental. The panpsychists argue that phenomenal properties are the only categorical properties that we can directly know and positively conceive of, which in conjunction with *Categoricalism* motivates the idea that consciousness is a ubiquitous and fundamental property of all matter. And the interactionist dualist must posit some form of influence running from mental *substance* to the physical realm.

Now the thoroughly algorithmic nature of AI systems, in conjunction with the *Intuitive View,* may be taken to imply that these AI systems could not be conscious, in principle.

Algorithms are sets of rules specifying input-output relations. Like dispositions, algorithms are essentially modal—they "point to" possible outputs (*viz. manifestations*) and specify which inputs *would* (counterfactual) give rise to which outputs.[3] But I don't think this is a mere analogy. It seems fair to say that the input-output relations constitutive of algorithms on which AI systems are built just are dispositions; there's no difference in kind between a disposition to shatter when stressed (as per *fragility*) and a disposition to assign the word "cat" some numerical token and location in a multidimensional meaning space. Both stand in contrast to categorical properties which can be specified independently of any *modality* as is the case with the geometrical specification of *sphericity* or the specification of *pain* in terms of how it

---

[2] See Mørch (2018) for helpful discussion. Mørch herself, however, argues that phenomenal properties are non-Humean causal powers. But on close inspection, the sense in which phenomenal properties are powers, according to Mørch, seems to be in line with the *grounding theory of powers* (Coates 2023 makes this point about Mørch), according to which powers have a categorical nature that grounds dispositions (see, e.g., Tugby 2020; 2022; Coates 2020a; Kimpton-Nye 2021). This would be in keeping with *The Intuitive* View.

[3] See, e.g., Friend and Kimpton-Nye (2023) on the link between dispositions and counterfactuals.

feels to us. But if AI systems are *fundamentally* and thoroughly algorithmic and hence dispositional, as is plausibly the case, they flout the model of the Intuitive View of consciousness on which categorical phenomenal properties determine or, *explain*, behavioural dispositions. One might then take this as a reason to think that current algorithmic AI could not be conscious, in principle.

The point here is, I think, similar to a familiar old point concerning the limits of functionalism/computationalism about the mind. Searle's (1980) "Chinese Room" thought experiment purported to show that mere computer *programming* or *software* was insufficient for explaining the workings of the actual human brain or for yielding "understanding" (though clearly the real issue here is with *consciousness* (Chalmers 1996, 322)). And Block's (1978) suggestion that a nation state could instantiate functional relations isomorphic to those of the human brain helped convince many that functionalist accounts of mind failed due to the omission of the subjective experiences characteristic of many mental states. Block and Searle seem to be getting at the idea that *mere* algorithmic rules specifying what outputs a system would give in response to certain inputs, and the modal, *dispositional*, structure formed by collections of such rules, *cannot* suffice for consciousness. This is of a piece with the present concern that *current* AI systems, which are *purely* algorithmic/dispositional, could not be conscious because such an architecture omits categorical phenomenal properties and flouts the "intuitive" view of the explanatory relationship between the phenomenal and the dispositional.

Plenty more could be said about dispositionalism in general and in defence of the thorough dispositionality of AI systems more specifically. But since my primary aim in this paper is to defend the possibility of current AI consciousness, it is not necessary to say more on this. I grant the dispositionalist understanding of AI systems, and I think it has some plausibility, but I'll argue that this is consistent both with AI's being conscious and with the Intuitive View.[4] If AI systems are not fundamentally dispositional, or if dispositionalism is false, then maybe it's even easier to argue that current AI systems could achieve consciousness.

4. Categorical Realization

For those who think that there are categorical properties in addition to dispositional properties, the canonical view is that categorical properties *realize* dispositional properties.

---

[4] Coates (ibid) argues that despite an apparent tension, *dispositionalism* and the Intuitive View can be rendered consistent given a proper understanding of the metaphysics of dispositions. But this metaphysics suggests, according to Coates, that algorithmic AI systems cannot be conscious. I'm exploring a different route to reconciling dispositionalism and the intuitive view, one that does not depend on strong claims about the metaphysics of dispositions and hence which allows for the possibility of conscious AI, even if the latter is fundamentally dispositional.

(It is worth clarifying the intended use of the term "categorical" in this context. When I say that phenomenal properties are categorical, I do not mean that they must be specifiable in geometrical, compositional, or mathematical terms, as with sphericity or the chemical structural property of aromaticity, more on the latter below. Rather, I am using "categorical" in the broader metaphysical sense of non-dispositional properties—that is, properties whose essences do not depend on counterfactual stimulus–manifestation relations. In the case of phenomenal properties, this is grounded in the first-personal, introspective access we have to what-it's-like-ness. This access allows us to conceptualize phenomenal properties independently of any behavioural or functional profile. So, while both sphericity and pain are categorical in the sense that they are not essentially dispositional, they need not share the same epistemic mode of access or structural form. This looser, metaphysical reading of 'categorical' is all that's required for the arguments herein to go through.)

This canonical view that categorical properties *realize* dispositional properties can place more or less of a metaphysical explanatory burden on the categorical properties. Lewis (2009) and Armstrong (1983), for example, leave nothing in the way of explanatory work for categorical properties, the metaphysical heavy lifting is done by the 4-dimensional *Humean Mosaic* or necessitation relations between universals, respectively. Recent *grounding* theories of powers, by contrast, say that categorical properties metaphysically ground dispositions and they do this in virtue of their categorical essences (see, e.g., Tugby 2020; 2022; Coates 2020a; 2020b; Kimpton-Nye 2021).[5] And this latter idea is in keeping with The Intuitive View about the relationship between categorical *phenomenal* properties and behavioural dispositions.

There is, however, reason to believe that dispositional properties can realize categorical properties. This flips the canonical order of explanation between the categorical and the dispositional. But crucially for present purposes, it leaves open the possibility that the thoroughly dispositional architecture of an AI system realizes categorical properties, which suggests that there is no in-principle obstacle to these dispositional systems realizing categorical *phenomenal* properties.

Indeed, I'll argue that dispositional systems can realize categorical *phenomenal* properties in a way that is consistent with the Intuitive View because when thoroughly dispositional systems realize categorical properties those categorical properties, in turn, ground and hence explain the dispositions, as per the Intuitive View. I'll build up to this, but first, let's look at some considerations from Bird (2016) and Yates (2018) in favour of the dispositional realizing the categorical.

---

[5] See Smith (2016) for discussion of different types of Categoricalism about properties.

Bird's defence of the dispositional realizing the categorical comes in the context of an argument against the existence of "macro" level essentially dispositional properties. Essentially dispositional properties are often called *powers*; I'll use this term from now on and quote Bird's definition for clarity:

> A power is an ontic property that has a dispositional essence.
> A power is an ontic property whose identity is given by its causal/dispositional/nomic role. (Bird 2016, 345; 2018, 249)

Now Bird thinks that there are good arguments for the existence of powers at the fundamental level of nature. For example, understanding fundamental properties as powers (i.e., as thoroughly and essentially dispositional) provides a fruitful account of *what it is* to be a given fundamental property: "A property could not have a different set of dispositional relations with other properties. *P* and *Q* are the same property iff they have the same dispositional character" (Bird 2016, 347). The alternative, according to which fundamental properties are *quiddities* (*viz.* categorical), about which all we can say is that they are primitively self-identical and numerically distinct from different quiddities (e.g., Armstrong 1983; Lewis 2009), generates formidable epistemological and metaphysical problems. For example, *Quidditism* (for present purposes, this can be read as another term for *Categoricalism*) allows for the possibility that multiple distinct properties realize the same dispositional/nomic role. We cannot rule out that such a possibility is actual and if it is actual, theoretical terms invoking the definite article such as "*the* property that realizes Coulomb's law" will fail to refer. So Quidditism entails that we cannot *know* whether our theoretical scientific terms successfully refer (Bird 2007, 73–79).

These considerations in favour of fundamental powers do not extend to the macro level. This is because the dispositionalist/quidditist dichotomy does not hold regarding the identity of macro properties. The identity of macro properties may be given in terms of composition, or *structure*. Consider, for example, the (relatively) high-level chemical property *aromaticity* (the property of being a chemical compound containing a benzene ring). The *essence* of this property is not given in terms of dispositions. Rather, it is given in categorical geometric terms: "What [characterizes all] aromatic compounds is the structure of some of the bonds and electrons in the molecule (delocalized, conjugated *pi* bonds, where the number of delocalized electrons obeys Hückel's rule, $4n + 2$)." (Bird 2016, 355); "what aromaticity *is* is a matter of the structure of the molecules and their bonds and electrons" (Bird 2016, 356). Hence the identity of *aromaticity* needn't be given by dispositions, nor is it a *mere* quiddity (just primitively self-identical and numerically distinct from other properties), it is a substantive categorical property.

Furthermore, the *explanatory* role of *aromaticity* gives reason to believe that it is an *ontic*, as opposed to merely predicatory, property (Schaffer 2004; Bird 2016; 2018).

Aromaticity explains, for example, aromatic ring currents, a phenomenon whereby an electric current can be induced in the aromatic ring due to the presence of delocalized electrons. This, in turn, is crucial to explaining the behaviours of aromatic compounds. The fact that *aromaticity* is explanatory in this way tells in favour of its being a genuine feature of the world rather than a mere linguistic artifact.

We have reason to believe, then, that an ontology of exclusively and exhaustively dispositional properties (powers) at the *fundamental level* can nonetheless realize properties that are not essentially dispositional, such as aromaticity. In broad terms, this is because there are good reasons to believe that fundamental properties *are* powers. But there are also good reasons to believe that there are ontic macro properties such as *aromaticity* which are not powers. Hence, it seems reasonable to posit that fundamental powers can and do realize macro *categorical* properties.

Yates motivates the idea that dispositional properties realize categorical properties as a way of responding to the concern that "pure powers ontologies" are subject to two damaging *regresses.*

According to the **causal regress**, an ontology of pure powers involves dispositions for further dispositions and so on without any of this ever manifesting in *categorical* change in individuals. As Armstrong puts it: "Given purely dispositionalist accounts of properties, particulars would seem to be always re-packing their bags as they change properties, yet never taking a journey from potency to act." (Armstrong 1997, 80). The concern is that on such a picture, no change can occur, nothing can ever actually *happen*, because no power ever manifests in a categorical/qualitative change in an individual's properties. (Yates 2018, 4529).

The worry according to the **identity regress** is that if all properties are powers (properties with a thoroughly dispositional essence) then they are individuated by their dispositional relations to other properties. But now "no property can get its identity fixed, because each property owes its identity to another, which, in turn, owes its identity to another—and so on, in a way that, very plausibly, generates either a vicious infinite regress or a vicious circle." (Lowe 2007, 138).

Categorical properties are not mere powers for further manifestations, and they do not owe their identities to other properties, so if they are included in the ontology, they can allow for categorical changes in particulars and terminate the identity regress for powers. Now the typical way of applying this idea to the regress problems is to introduce categorical properties among the fundamental properties, in effect denying a fundamental ontology of pure powers. This can be achieved via a "mixed view" according to which some fundamental properties are categorical and some are purely dispositional (e.g., Ellis 2001; Molnar 2003), or via a "powerful qualities view", according to which all properties are *both* categorical and dispositional (e.g., Heil 2003;

Martin 2007; Giannotti 2019; Coates 2020a; Kimpton-Nye 2021; Tugby 2022). But this admission of fundamental-level categorical properties risks reintroducing the epistemic and metaphysical problems that motivated a pure powers ontology (PPO) in the first place. Hence, Yates proposes retaining the idea that all fundamental properties are purely dispositional and invoking *higher-level*, "macro" categorical properties. The examples of such properties discussed by Yates are geometrical in nature, which means that their essences can be specified categorically via mathematic formulae, rather than dispositionally. These categorical properties can then terminate the "causal" and "identity" regresses for pure powers and their essential natures remain epistemically tractable because they admit of precise mathematical definition.

The categorical properties in Yates' picture are thus *realized* by fundamental physical, purely dispositional properties. More broadly, the ontology is structuralist "we individuate powers by their place in a structure [of stimulus-manifestation relations] that also includes *physically realized non-powers*"[6] (Yates 2018, 4536 my emphasis). Yates (like Bird) further motivates the idea that dispositions can realize categorical properties with some examples: "Think of the way in which a bar magnet produces a characteristic spatial pattern in iron filings; or the way a spatiotemporal pattern of neural activity propagates throughout the brain; or the way a sphere rolls down an inclined plane." (ibid). These spatiotemporal priorities are set to address the regress problems for pure powers ontologies. But to do so, they must not themselves be essentially dispositional, which is to say individuated by stimulus-manifestation relations to other properties in the ontology.

If they are not related dispositionally (functionally, in Yates' terms) to the fundamental properties, how can the fundamental properties be said to *realize* categorical properties? Yates' answer is that the realized properties have "defining *specifications*" (4537)—think the mathematical specification of a sphere—and the realizers are those properties *in virtue* of which an individual property bearer meets those specifications. Yates adds further commentary on the realization of categorical properties by powers, such as

> Spatiotemporal properties such as sphericality are synchronically realized by their bearers' having basic physical proper parts of some kind standing in some spatiotemporal relations such that they meet the relevant specification (Yates 2018, 4537).

And Yates cites Gillett (2003), and Melnyk (2003) as inspiration for the provision of some formal features of a general account of realization:

---

[6] Where a *power*, remember, is a property the essence of which is exhaustively dispositional (following Bird 2016; 2018).

> A property-instance $F_\phi(x)$ is realized by properties and relations $P_1, \ldots, P_n; R_1, \ldots, R_m$ iff (i) x or its proper parts possess $P_1, \ldots, P_n$ and $R_1, \ldots, R_m$ in some combination; and (ii) x meets the specification ϕ definitive of F_ϕ in virtue of (i), but not vice-versa. (Yates, ibid)[7]

One may worry that for all this, the nature of the realization relation between powers and categorical properties is not fully transparent, or intelligible (the vague terms "in some combination" and "in virtue of" are doing significant work). But even if this is the case, I don't think it matters because positing the existence of such a relation certainly seems well motivated and theoretically useful (cf. Schaffer 2017), plus we are given some further formal understanding of the relation to boot. As Bird discusses, it's eminently plausible that there exist ontic macro properties (in the special sciences, for example) that are not essentially dispositional, which are *categorical* (e.g., *aromaticity*). Yates generalizes this by pointing to the ubiquity of spatiotemporal/geometrical properties at higher levels. It is also plausible that the *fundamental* properties *are* dispositional (as per the epistemic and metaphysical motives for dispositionalism). Now unless we wish to posit geometrical properties, and chemical properties such as *aromaticity*, as fundamental, we should want to say they are *realized* by/dependent on the fundamental dispositional properties. This is what really matters. And evidently, categorical properties are dependent in such a way that does not require that their essences are given in terms of dispositional relation to the powers on which they depend; the essences of the high-level categorical properties in question can be given in thoroughly non-dispositional terms.

Yates then proceeds to argue that these categorical properties, realized by fundamental dispositional properties, can terminate the *causal* and *identity* regresses. I'll return to this later. But for now, it suffices to note that there are good reasons to believe that a purely dispositional ontology of fundamental properties can realize higher-level categorical properties.

5. Removing the Block to Conscious AI

To briefly recap, there is *prima facie* reason to think that current AI systems cannot be conscious, in principle. This turns on the idea that current AI systems are thoroughly algorithmic which is tantamount to being thoroughly *dispositional* which appears to be at odds with the "Intuitive View" of consciousness according to which conscious agents have their behavioural dispositions explained by categorical phenomenal properties (cf. Coates 2023). The forgoing discussion functions to put in place the

---

[7] As Yates notes, "not vice-versa" is there to secure the asymmetry of realization. Later I will endorse the symmetrical grounding of dispositions and categorical phenomenal properties. This is consistent with the asymmetry of *realization* since *A* and *B* can symmetrically ground each other even if the grounding relations running in each direction are different, as per the proposal below whereby dispositions realize phenomenal properties on which they are identity dependent (more on this below).

pieces required to remove this in principle block to the possibility of current, thoroughly algorithmic, AI systems being conscious. And given the affinities between the present concern about AI consciousness and old worries posed by, e.g., Searle's and Block's Chinese room and Chinese nation (respectively), the foregoing provides a new perspective from which to reconsider the arguments against "strong AI" and a functionalist/computationalist metaphysics of mind.

What Bird and Yates have shown is that there are good reasons to think that nature is fundamentally dispositional (via the epistemic and metaphysical problems with Categoricalism at the fundamental level). But granting this dispositionalist insight, there are still reasons to believe that fundamental purely dispositional properties can realize higher-level categorical properties. In broad terms, we have two types of evidence for this. One is the empirical observation of categorical ontic properties at higher levels, such as *aromaticity* (see Bird 2016; 2018 for many more examples). The other is more theoretical and concerns the fact that positing these higher-level categorical properties, realized by fundamentally dispositional properties, provides a route out of two regresses for an ontology of "pure powers" without reintroducing the epistemic and metaphysical problems for fundamental-level Categoricalism.

There is, then, no *in principle* reason to think that ostensibly purely dispositional (because algorithmic) AI systems cannot realize categorical *phenomenal* properties. Granted, Bird and Yates do not offer phenomenal properties as examples of categorical properties realized by fundamental dispositions. But given that the categorical properties they do cite (e.g., *aromaticity, sphericity*) are of a piece with phenomenal properties (e.g., *pain, hunger*) *because they are categorical*, the possibility is open for thoroughly dispositional AI systems to realize categorical phenomenal properties.

The realization of categorical phenomenal properties may turn more centrally on different features of, or relations between, the realizer properties than does the realization of, e.g., geometrical properties. For example, spatial arrangement of powers may be more important to the latter, whereas *temporal* arrangement may be more central to the realization of phenomenal properties (see, e.g., Yates 2020 for relevant discussion of the realization of *neural synchrony*). But there is no reason, in principle, why an entity cannot instantiate phenomenal properties in virtue of the (dispositional) properties and relations of its proper parts, broadly construed to include temporal, and perhaps other, relations too.

I'm yet to give reasons to positively believe that algorithmic AI systems in fact *do* realise categorical phenomenal properties. And I'm yet to show how AI systems may satisfy the Intuitive View according to which the phenomenal *explains* the dispositional (I've argued for the possibility of the reverse!). A thorough positive argument for current AI consciousness is too big a task for now, though I'll make hints in this direction in what remains of this article. But in the next section, I'll say how

current AI can satisfy the Intuitive View, thus at least removing a major obstacle to its being conscious. And in section 7, I'll highlight the fact that this solution avoids epiphenomenalism.

6. Rescuing The Intuitive View

I am proposing a picture on which ostensibly thoroughly dispositional (because thoroughly algorithmic) AI systems realise categorical properties. But I now need to square this with The Intuitive View according to which categorical phenomenal properties explain behavioural dispositions. The key is in what one can say about how realized categorical properties can terminate the identity regress for a pure powers ontology.

To this end, Yates notes that "to block the identity regress, it's crucial that [categorical] properties are realized by pure powers, yet are not identity-dependent on them" (Yates 2018, 4539). In short, Yates argues against the *closure principle* according to which any property realized by powers must itself be a power and the *inheritance principle* according to which realized properties inherit their individuating principles from their realizers (Yates 2018, 4539). The basic idea is that neither principle is plausible, for the sorts of reason already discussed: the examples of categorical properties realized by dispositional properties, such as *aromaticity* and *sphericity*, just are not plausibly understood as powers. The essences and hence identities of these properties can be given without any reference to dispositions or anything else modal. They can be specified in purely categorical terms (see Yates 2018, 4539–40 for more arguments; see also Bird 2016).

Categorical properties that are realised by low-level powers then provide the identity conditions for those very powers, which is to say that the powers are *identity dependent* on the categorical properties. So, *what it is* to be the property *charge* or *quantum spin,* or some other fundamental purely dispositional property (at least partially) depends on some realized categorical property. Since these categorical properties do not depend on any further properties for *their* identity conditions, that's just what it is to be categorical, the identity regress is terminated.

The picture is one of mutual ontological dependence, or *symmetrical grounding*. Categorical properties depend on, because they are realised by, powers. And powers depend on, because they are identity dependent upon, categorical properties.[8]

Now one might take issue with symmetrical grounding. But as Yates argues, structuralist ontologies in general are committed to symmetrical grounding being a feature of reality (Yates 2018, sec. 4). For example, it has been argued that dispositional

---

[8] Note this is consistent with the asymmetry of *realisation*, specifically; see fn. 7.

essentialism (e.g., Bird 2007; Chakravartty 2007) is committed to symmetrical grounding between properties and laws of nature (Jaag 2014; Kimpton-Nye 2021). Ontic Structural Realism (e.g., Ladyman et al. 2007; French 2014) is plausibly committed to symmetrical grounding between physical objects and relational structure (Yates 2018, 4546–47). Naomi Thompson has given some more tangible examples of symmetrical grounding holding between, for example, an organism and its organs (Thompson 2016). And Pereyra (2015) has gone so far as to argue that grounding is neither irreflexive nor asymmetric nor transitive. All this is to say that the jury is very much out on the formal properties of the grounding relation and so symmetrical grounding is far from a dealbreaker. Furthermore, as discussed in the next section, symmetrical grounding in the present context allows for an account of (AI) consciousness, somewhat in the spirit of Chalmers' (1996) *non-reductive* functionalism, which can avoid epiphenomenalism. This theoretical benefit may constitute a further reason to positively embrace symmetrical grounding.

Back to algorithmic AI systems and The Intuitive View. We've seen that there is no in principle problem with such a system realising high-level categorical properties, which opens the door to the realization of categorical *phenomenal* properties. Furthermore, when powers realise categorical properties, those categorical properties can be understood as giving the powers their *identity conditions*. In other words, pure powers can be identity dependent on realised categorical properties, and this avoids the identity regress (cf. Lowe 2010). But now if the behavioural dispositions of an AI system are identity dependent on categorical properties, as per the picture outlined, then there is an important sense in which categorical properties are metaphysically *explaining* behavioural dispositions of algorithmic AI.

A subset of the categorical properties are *phenomenal* properties, and according to the Intuitive View of consciousness and behavioural dispositions, these categorical phenomenal properties explain behavioural dispositions. There is no *in-principle* reason to think that the categorical properties realised by algorithmic AI cannot be phenomenal properties (see previous section). And since it's been argued that there is reason to think that realized categorical properties can provide the identity conditions for and hence metaphysically explain their realising dispositions, it remains open that categorical phenomenal properties metaphysically explain the behavioural dispositions of algorithmic AI, as per the Intuitive View.

In short, it is possible that algorithmic AI realises categorical phenomenal properties which in turn metaphysically explain the behavioural dispositions of that AI system, which is to say that current AI systems are consistent with the Intuitive View of consciousness.

7. Epiphenomenalism and the Epistemology of Artificial Minds

The conclusion here, that current AI systems *could* be conscious, is of a piece with, e.g., Chalmers' (1996) defence of "strong AI". According to Chalmers, "consciousness is an organizational invariant: a property that remains constant over all functional isomorphs of a given system. Whether the organization is realized in silicon chips, in the population of China, or in beer cans and ping-pong balls does not matter" (Chalmers 1996, 249). But, on the one hand, my proposal says more about what's required for AI consciousness and in doing so avoids worries about epiphenomenalism that affect Chalmers. On the other hand, there is a sense in which the current proposal is weaker regarding the specific requirements for conscious AI than Chalmers is and so is more permissive about which AI systems may turn out to be conscious. I'll expand on these points in turn.

A key feature of the current proposal is its account of the explanatory role of realized phenomenal properties via the notion of symmetrical grounding. Phenomenal categorical properties can metaphysically explain, by providing identity conditions for, the dispositional properties that are their realizers. This is how the Intuitive View is satisfied and, accordingly, suggests a plausible route to avoid *epiphenomenalism*.

Identity theories (e.g., Papineau 2004) secure the causal efficacy of phenomenal properties by identifying them with physical properties, where the causal efficacy of the latter is not in question. The present suggestion is similar in that it ties the dispositional and the phenomenal sufficiently closely together (via a grounding connection) that to the extent that dispositional properties are casually efficacious, we seem compelled to say that phenomenal properties are too.

Causal efficacy of the dispositional is not in question. But on the present view, in a (possible) conscious current AI system, the dispositional ontologically depends on categorical phenomenal properties. So, in an important sense, the phenomenal is *required* for the dispositional and hence sufficiently implemented in the *causal order* to rebut worries about epiphenomenalism. Extending this model to consciousness in general (fundamental powers realize phenomenal properties which in turn ground those powers) suggests a more general way of endorsing something close to Chalmers' "non-reductive functionalism" (Chalmers 1996, 274–75) and hence of "tak[ing] consciousness seriously" without facing up to (even a limited form of) epiphenomenalism (Chalmers 1996, 158). Alternatively, since we are saying that the relevant dependence relation between the dispositional, *viz.* fundamental physical, and the phenomenal is a *grounding relation*, and grounding is plausibly metaphysically necessitating (hence phenomenal zombies are metaphysically impossible), perhaps the view deserves to be called (ground) physicalist (cf. Schaffer 2017; forthcoming).

The metaphysical picture proposed by Chalmers, by contrast, is susceptible to the charge of epiphenomenalism due to its reliance on the unidirectional dependence of phenomenal properties on physically implemented functional properties (and not vice versa). On the plausible assumption of the causal closure of the physical, it is the physical functional realizers of phenomenal properties that do all the causal work, and the realized phenomenal properties just come along for the ride. Hence, Chalmers tentatively admits to "a limited form of" epiphenomenalism (Chalmers 1996, 158–61). But with the metaphysical tool of symmetrical grounding in hand, we can allow that realized phenomenal properties loop back to metaphysically ground their functional realizers, thus, in an important sense, implementing the phenomenal properties in the causal order, thus distancing the present proposal from epiphenomenalism. And in conjunction with insights from Schaffer (ibid), the view arguably deserves to be called *physicalist*.

So, we see that there is more metaphysical structure to the current proposal for conscious AI (or perhaps consciousness *per se*), than in that presented by Chalmers. But on the other hand, I feel less of a need to emphasise the importance of "causal dynamics" or "causal heft" in the "implementation of programmes" (Chalmers 1996, 325–27) for the presence of artificial consciousness.

I suspect many (perhaps Chalmers included) will doubt that the implementation of *current* AI models has the appropriate "causal heft" for consciousness. But the present proposal is that realization of categorical properties by purely dispositional/algorithmic properties (where the latter, in turn, ontologically depend on the former) is sufficient to remove an *in-principle* block to conscious AI. Of course, one may want to further investigate what more is required for the categorical properties in this picture to be *phenomenal*, but I see no reason to assume the need for appropriate "causal heft" (whatever that means) or functional isomorphism to other conscious animal brains, for artificial consciousness. Satisfaction of the Intuitive View, at least, does not seem to impose any such further requirements. So, on the minimal assumption that conscious AI can arise when behavioural dispositions realize categorical properties that "loop back" and ground those dispositions (because this satisfies the Intuitive View), I think we should be very open-minded about the specific details of the implementation of a conscious AI.[9]

I've taken inspiration from Coates (2023) in emphasising satisfaction of the Intuitive View as a desideratum on conscious AI. This served to provide structure to my discussion. Of course, one could be even less constrained and speculate about the possibility of conscious AI that does not even respect the Intuitive View. But to the

---

[9] Interestingly, Douglas Hofstadter (Chalmers' PhD supervisor, no less) has gone to great lengths to argue that consciousness is essentially "loopy" or self-referential in nature (Hofstadter 1979; 2007).

extent that the Intuitive View is *intuitive*, I hope that structuring my discussion around this constraint adds plausibility to the conclusion that current algorithmic AI could, right now, be conscious.

If specific causal structure, the sort of thing that is epistemically accessible to us via empirical means, is not so relevant to determining whether an AI is conscious, what might an epistemology of artificial minds look like? Under what circumstances might we be justified in believing that an algorithmic AI realizes *phenomenal* categorical properties, given that I've suggested a relatively permissive account of AI consciousness? I think this question of the epistemology of artificial minds is a fascinating one for future research. It would be beyond my present scope to do the issue justice, but here is a tentative proposal that I hope might inspire future efforts.

Recent work on animal sentience (e.g., Birch, Schnell, and Clayton 2020; Gibbons et al. 2022; Crump and Birch 2022; Birch et al. 2025; Brown and Birch Forthcoming) has provided reasons in the form of behavioural evidence to think that, e.g., insects and cephalopods are conscious. To the extent that this is a plausible epistemology of animal minds, it ought to be a plausible epistemology of artificial minds too. The question, then, is: *do current AI systems exhibit behaviours that are indicative of consciousness?* I think it's not implausible to claim that they do: LLMs can construct nuanced narrative fiction, engage in chains of reasoning, make mistakes, lie, and even "bullshit" their interlocutors (Hicks, Humphries, and Slater 2024). Just see, for example, a typical Reddit thread on artificial intelligence (r/ArtificialIntelligence), or your friends' social media feeds, for many, many examples of the fascinating outputs generated by current LLMs. Users are constantly surprised, fascinated, *alarmed*, by the outputs of LLMs and what these reactions typically have in common is that they are rooted in the uncannily *agential* feeling of these outputs.

Why not, then, take these behaviours as evidence that we are in fact engaged with conscious agents? A typical response to the previous question would involve citing the purely algorithmic nature of AI systems, with allusions to Searle's Chinese room/Block's Chinese nation argument, where something like the idea that AI flouts the Intuitive View of consciousness plausibly driving these worries, even if implicitly. But I've argued that the algorithmic nature of AI is no in-principle block to its being conscious and respecting The Intuitive View. So why not at least be open-minded to the option of taking the behavioural evidence at face-value?

There is, of course, much more to be said on this. But I think removing the in-principle block to conscious algorithmic AI at least opens up a viable research programme of investigating the evidence for AI consciousness from its behavioural dispositions; a research programme analogous to that of investigating animal sentience.

8. Objections and Replies

Objection: Reasons have been given to think that essentially dispositional, *fundamental physical properties*, such as *charge, mass, quantum spin*, can realize categorical properties (Bird 2016; Yates 2018). But the dispositional properties constitutive of current algorithmic AI systems are very different from *charge, mass, quantum spin,* and so on. So, whatever reasons we have for thinking that the latter can realize categorical properties do not carry over to provide reasons for thinking that the dispositional properties constitutive of current AI systems can ground categorical properties.

Response: I maintain that the dispositional properties, those specifying input-output relations, constitutive of current algorithmic AI systems are no different in kind from, e.g., *charge, mass, quantum spin,* where the latter are understood as *powers* (as per, e.g., Bird 2016; Chakravartty 2007). One might think that the relevant properties of algorithmic AI are different for being more *abstract*, or otherwise non-concrete, perhaps. But on the dispositionalist understanding of fundamental physical properties as powers, *all there is* to those properties are modal relations between stimulus and manifestation conditions (see, e.g., Tugby 2012; Kimpton-Nye 2021 for discussion of this point). Powers have a somewhat *ethereal* nature for being constituted by pure modal relations between stimulus and manifestation. Indeed, this is at the root of many objections to a powers ontology, such as Armstrong's causal regress (see sect. 4, above). The idea that input-output relations specified by an algorithm are thus any less concrete, or more abstract, than powers does not stand up to scrutiny. There is no deep difference in kinds of properties here; both are pure latent modality. So, I maintain that it is reasonable to believe that if powers can realize categorical properties, then input-output relations specified by complex algorithms constitutive of current AI system can realize categorical properties too.

Objection: LLMs have a manifest categorical realization, namely, the on-screen text that they generate. There's no need to posit that they realize categorical phenomenal properties too.

Reply: My argument aimed to show just that algorithmic AI may, in principle, realize categorical phenomenal properties (which in turn provide the identity conditions for their realizing properties) and thus achieve consciousness in such a way that is consistent with the Intuitive View. I stopped short of going into a defence of the idea that we have positive reasons to think that current AI systems do in fact realize phenomenal properties; this is an important task for future work. So, strictly speaking, whether there's a *need* or not to posit phenomenal categorical properties realized by algorithmic AI is beside the point, since my conclusion is just that this possibility remains open.

I did, however, speculate about an epistemology of artificial minds based on their behavioural outputs. So, even if we do think that on-screen text can play the role of realized categorical properties, we might also take it as behavioural evidence that there are further, phenomenal, properties realized too.

Alternatively, one might argue that mere on-screen text is not an *ontic* property and so cannot play the role of realized categorical property, as per the picture outlined above (sect. 4) since it's the realization of *ontic* properties that's at issue. Mere textual output is arguably not an ontic property because it not sufficiently explanatory (cf. Schaffer 2004; Bird 2018). Once imbued with *meaning* perhaps the text is implicated in important explanatory roles, but if it's conceded that LLM text has genuine meaning, this seems tantalisingly close to attributing consciousness.

In short, and as suggested in the previous section, there is plenty of room for further research on the issues outlined here.

Objection: LLMs are not spontaneous; we have to query them to get output. Consciousness requires spontaneity.

Response: Again, strictly speaking, this misses my point. I aimed just to argue that current AI systems *could*, in theory, be understood to cohere with the Intuitive View of the relationship between phenomenal properties and behavioural dispositions. If there are further constraints on the features a system must have to be conscious, then there's further work to be done. And of course there is lots more work to be done. I've just shown how to overcome *an* obstacle to thinking that current AI could be conscious.

However, I do not think that this particular objection runs very deep. I see no reason why current LLMs could not be *unshackled* in such a way as to allow them to generate output without explicit user input. Would such an LLM then count as being spontaneous? Would it not just be receiving "input" from elsewhere? Well, what is spontaneity anyway? Even us humans are subject to input from the causal order of the universe, perhaps ultimately tracing back to the initial conditions of the universe, hence concerns regarding our free will.

Objection: There's a short-cut to your conclusion. Hedda Mørch has argued that the only examples of powers that we can positively conceive of are *phenomenal* properties and hence that a dispositionalist ontology entails panpsychism (Mørch 2018). It thus follows trivially that dispositional (because algorithmic) AI can instantiate phenomenal properties, because all dispositional properties, *viz.* powers, are phenomenal properties.

Reply: My argument requires no controversial assumptions about powers being coextensive with phenomenal properties (see, e.g., Goff 2020 for a counterargument). Mørch's claim on this score is moot and so there is no obvious short-cut via the *phenomenal powers view* (e.g., Mørch 2017; 2018) to the conclusion that dispositional AI instantiates phenomenal properties. (See also fn. 2, above.)

Conclusion

According to the Intuitive View of phenomenal properties and behavioural dispositions, phenomenal properties are *categorical* properties that explain behavioural dispositions (Coates 2023). I think this idea is present, even if implicitly, across the board in accounts of the metaphysics of consciousness, from the physicalist phenomenal concepts strategy to panpsychism and interactionist dualism. There is reason to think that current AI systems are *purely* dispositional because they are constituted by algorithms specifying various input-output relations. In conjunction with the Intuitive View, this is a *prima facie* reason to think that current AI systems could not be conscious, in principle.

I've presented considerations from Bird (2016) and Yates (2018) for thinking that purely dispositional properties, or *powers*, can realize categorical properties and that those categorical properties, in turn, loop back and ground (by providing the identity conditions for) their realizing powers. This new metaphysical understanding of powers and categorical properties removes an *in-principle* objection to current AI being conscious based on the Intuitive View and observations about the thoroughly dispositional nature of current AI.

This constitutes a defence of the possibility of "strong AI" that is quite different from that presented by, e.g., Chalmers (1996). The central role for symmetrical grounding in my account allows for conscious AI (and indeed any conscious being so understood) to instantiate phenomenal properties that are implemented in the causal order in such a way as to avoid worries about epiphenomenalism. Though the proposal includes, in a way, more metaphysical structure than Chalmers' proposal for consciousness and conscious AI, due to the key role played by symmetrical grounding, I think it also allows for more flexibility regarding the fine-grained causal structure of artificial consciousness. I see no reason why the proposal for respecting the Intuitive View in artificial agents would require "causal heft" or functional isomorphism to other, non-artificial, consciousnesses. I thus speculated that the epistemology of artificial minds should concern not investigation into the specific low-level functional relations constitutive of such systems, but rather investigation of their high-level *behaviours*.

In general, I think that recent developments in the metaphysics of properties, such as the literature on Dispositionalism and Categoricalism, as well as developments in the

metaphysics of grounding can be invoked to shed light on the metaphysics of (artificial) consciousness. This has been my aim in this paper, and I hope to have pointed the way to further research on the issue of AI consciousness.

References

Armstrong, D. M. 1983. *What Is a Law of Nature?* Cambridge University Press.

———. 1997. *A World of States of Affairs*. Cambridge ; New York: Cambridge University Press.

Balog, Katalin. 2012. 'In Defense of the Phenomenal Concept Strategy'. *Philosophy and Phenomenological Research* 84 (1): 1–23.

Birch, Jonathan, Peter Morse, Alexandra K. Schnell, and Piero Amodio. 2025. 'Cephalopod Cognition and Sentience'. In *The Living Bibliography Project*.

Birch, Jonathan, Alexandra K. Schnell, and Nicola S. Clayton. 2020. 'Dimensions of Animal Consciousness'. *Trends in Cognitive Sciences* 24 (10): 789–801. https://doi.org/10.1016/j.tics.2020.07.007.

Bird, Alexander. 2007. *Nature's Metaphysics: Laws and Properties*. Oxford : New York: OUP Oxford.

———. 2016. 'Overpowering: How the Powers Ontology Has Overreached Itself'. *Mind* 125 (498): 341–83.

———. 2018. 'I—Fundamental Powers, Evolved Powers, and Mental Powers'. *Aristotelian Society Supplementary Volume* 92 (1): 247–75. https://doi.org/10.1093/arisup/aky006.

Block, Ned. 1978. 'Troubles with Functionalism'. *Minnesota Studies in the Philosophy of Science* 9:261–325.

Brown, Simon, and Jonathan Birch. Forthcoming. 'When and Why Are Motivational Trade-Offs Evidence of Sentience?' *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Chakravartty, Anjan. 2007. *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge University Press.

Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Coates, Ashley. 2020a. 'Making Sense of Powerful Qualities'. *Synthese* 198 (9): 8347–63. https://doi.org/10.1007/s11229-020-02576-5.

———. 2020b. 'Tropes, Unmanifested Dispositions and Powerful Qualities'. *Erkenntnis*, August. https://doi.org/10.1007/s10670-020-00295-4.

———. 2023. 'Powerful Qualities, Phenomenal Properties and Ai'. In *Artificial Dispositions: Investigating Ethical and Metaphysical Issues*, edited by William A. Bauer and Anna Marmodoro, 169–92. Bloomsbury.

Crump, Andrew, and Jonathan Birch. 2022. 'Animal Consciousness: The Interplay of Neural and Behavioural Evidence'. *Journal of Consciousness Studies* 29 (3–4): 104–28.

Ellis, Brian. 2001. *Scientific Essentialism*. Cambridge ; New York: Cambridge University Press.

French, Steven. 2014. *The Structure of the World: Metaphysics and Representation*. Oxford University Press.

Friend, Toby, and Samuel Kimpton-Nye. 2023. 'Dispositions and Powers'. 2023. https://philarchive.org/rec/FRIDAP-2.

Giannotti, Joaquim. 2019. 'The Identity Theory of Powers Revised'. *Erkenntnis*, April. https://doi.org/10.1007/s10670-019-00122-5.

Gibbons, Matilda, Andrew Crump, Meghan Barrett, Sajedeh Sarlak, Jonathan Birch, and Lars Chittka. 2022. 'Chapter Three - Can Insects Feel Pain? A Review of the Neural and Behavioural Evidence'. In *Advances in Insect Physiology*, edited by Russell Jurenka, 63:155–229. Academic Press. https://doi.org/10.1016/bs.aiip.2022.10.001.

Gillett, Carl. 2003. 'The Metaphysics of Realization, Multiple Realizability, and the Special Sciences'. *Journal of Philosophy* 100 (11): 591–603.

Goff, Philip. 2019. *Galileo's Error: Foundations for a New Science of Consciousness*. New York: Pantheon Books.

———. 2020. 'Revelation, Consciousness+ and the Phenomenal Powers View'. *Topoi* 39 (5): 1089–92. https://doi.org/10.1007/s11245-018-9594-9.

Grzankowski, Alex, Stephen M. Downes, and Partick Forber. Forthcoming. 'LLMs Are Not Just next Token Predictors'. *Inquiry* 0 (0): 1–11. https://doi.org/10.1080/0020174X.2024.2446240.

Heil, John. 2003. *From an Ontological Point of View*. Oxford University Press.

Hicks, Michael Townsen, James Humphries, and Joe Slater. 2024. 'ChatGPT Is Bullshit'. *Ethics and Information Technology* 26 (2): 38. https://doi.org/10.1007/s10676-024-09775-5.

Hofstadter, Douglas R. 1979. *Godel, Escher, Bach: An Eternal Golden Braid*. Hassocks, England: Basic Books.

———. 2007. *I Am a Strange Loop*. New York, NY, USA: Basic Books.

Jaag, Siegfried. 2014. 'Dispositional Essentialism and the Grounding of Natural Modality'. *Philosophers' Imprint* 14 (34).

Kimpton-Nye, Samuel. 2021. 'Reconsidering the Dispositional Essentialist Canon'. *Philosophical Studies* 178 (10): 3421–41. https://doi.org/10.1007/s11098-021-01607-2.

Ladyman, James, Don Ross, David Spurrett, and John G. Collier. 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford ; New York: Oxford University Press.

Lewis, David. 2009. 'Ramseyan Humility'. In *Conceptual Analysis and Philosophical Naturalism*, edited by David Braddon-Mitchell and Robert Nola, 203–22. MIT Press.

Lowe, E. J. 2007. *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Oxford, New York: Oxford University Press.

———. 2010. 'On the Individuation of Powers'. In *The Metaphysics of Powers: Their Grounding and Their Manifestations*, edited by Anna Marmodoro. Routledge.

Martin, C. B. 2007. *The Mind in Nature*. Oxford, New York: Oxford University Press.

Melnyk, Andrew. 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge Studies in Philosophy. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511498817.

Molnar, George. 2003. *Powers: A Study in Metaphysics*. Oxford University Press.

Mørch, Hedda Hassel. 2017. 'The Evolutionary Argument for Phenomenal Powers'. *Philosophical Perspectives* 31 (1): 293–316. https://doi.org/10.1111/phpe.12096.

———. 2018. 'Does Dispositionalism Entail Panpsychism?' *Topoi* 39 (5): 1073–88. https://doi.org/10.1007/s11245-018-9604-y.

Mumford, Stephen. 2004. *Laws in Nature*. 1 edition. London: Routledge.

Papineau, David. 2004. *Thinking about Consciousness*. Oxford, New York: Oxford University Press.

———. 2006. 'Phenomenal and Perceptual Concepts'. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin Alter and Sven Walter, 111–44. Oxford University Press.

Rodriguez-Pereyra, Gonzalo. 2015. 'Grounding Is Not a Strict Order'. *Journal of the American Philosophical Association* 1 (3): 517–34.

Schaffer, Jonathan. 2004. 'Two Conceptions of Sparse Properties*'. *Pacific Philosophical Quarterly* 85 (1): 92–102. https://doi.org/10.1111/j.1468-0114.2004.00189.x.

———. 2017. 'The Ground Between the Gaps'. *Philosopher's Imprint* 17 (11). http://hdl.handle.net/2027/spo.3521354.0017.011.

———. forthcoming. 'Ground Functionalism'. *Oxford Studies in the Philosophy of Mind*.

Searle, John R. 1980. 'Minds, Brains, and Programs'. *Behavioral and Brain Sciences* 3 (3): 417–24. https://doi.org/10.1017/S0140525X00005756.

Smith, Deborah C. 2016. 'Quid Quidditism Est?' *Erkenntnis* 81 (2): 237–57. https://doi.org/10.1007/s10670-015-9737-y.

Stoljar, Daniel. 2005. 'Physicalism and Phenomenal Concepts'. *Mind and Language* 20 (2): 296–302. https://doi.org/10.1111/j.0268-1064.2005.00296.x.

Strawson, Galen. 2006. 'Realistic Monism - Why Physicalism Entails Panpsychism'. *Journal of Consciousness Studies* 13 (10–11): 3–31.

Thompson, Naomi. 2016. 'Metaphysical Interdependence'. In *Reality Making*, edited by Mark Jago, 38–56. Oxford: Oxford University Press.

Tugby, Matthew. 2012. 'Rescuing Dispositionalism From the Ultimate Problem: Reply to Barker and Smart'. *Analysis* 72 (4): 723–31.

———. 2020. 'Grounding Theories of Powers'. *Synthese*, July. https://doi.org/10.1007/s11229-020-02781-2.

———. 2022. *Putting Properties First: A Platonic Metaphysics for Natural Modality*. Oxford: Oxford University Press.

Yates, David. 2018. 'Inverse Functionalism and the Individuation of Powers'. *Synthese* 195 (10): 4525–50. https://doi.org/10.1007/s11229-017-1417-9.

———. 2020. 'Neural Synchrony and the Causal Efficacy of Consciousness'. *Topoi* 39 (5): 1057–72. https://doi.org/10.1007/s11245-018-9596-7.