To appear in *Mental Powers* (eds. Marmodoro & Grasso), *Topoi* 2018. Published version is free to view at https://rdcu.be/7wV6.

Neural Synchrony and the Causal Efficacy of Consciousness

The purpose of this paper is to address a well-known dilemma for physicalism. If mental properties are type identical to physical properties, then their causal efficacy is secure, but at the cost of ruling out mentality in creatures very different to ourselves. On the other hand, if mental properties are multiply realizable, then all kinds of creatures can instantiate them, but then they seem to be causally redundant. The *causal exclusion problem* depends on the widely held principle that realized properties inherit their causal powers from their realizers. While this principle holds for functional realization, it fails on a broader notion of realization that permits the realization of complex qualitative properties such as spatial and temporal patterns. Such properties are best seen as dependent powerful qualities, which have their causal roles in virtue of being the qualities they are, and do not inherit powers from their realizers. Recent studies have identified one such property—*neural synchrony*—as a correlate of consciousness. If synchrony is also partially *constitutive* of consciousness, then phenomenal properties are both multiply realizable and causally novel. I outline a version of representationalism about consciousness on which this constitution claim holds.

1. Identity or Realization: A Dilemma for Physicalism

Physicalists have a perennial problem squaring metaphysics of mind with mental causation. Causal efficacy seems to require the type identity theory, on pain of violating the causal closure of the physical. But the identity theory, it is widely assumed, entails that only creatures with brains like ours get to have minds. On the other hand, if we suppose that mental properties are not physical but physically realized, then physically different creatures get to have minds, but realized properties seem to be causally excluded by their realizers.

To appear in *Mental Powers* (eds. Marmodoro & Grasso), *Topoi* 2018. Published version is free to view at https://rdcu.be/7wV6.

Consider the following seemingly inconsistent set of propositions, each of which we have good reason to believe, and which jointly give rise to a causal exclusion problem for consciousness:[1]

1. Phenomenal properties cause physical effects[2]
2. The physical domain is causally closed
3. The effects of mental properties are not generally overdetermined
4. Phenomenal properties are not identical to physical properties

By (1), properties like the phenomenal character of a colour experience or a pain bestow causal powers to bring about physical effects, for instance verbal and non-verbal behaviour. By (2), whatever physical effects such properties cause have fully sufficient physical causes. If, as (3) states, mental causation is not causal overdetermination, then it seems we must identify phenomenal and physical properties. But by (4), the type identity theory is false.

There are of course many ways for physicalists to respond. Some bite the bullet and deny that phenomenal properties do any genuine causal work. The challenge then is to give an account of causal explanations featuring such properties that doesn't require causal efficacy.[3] Others appeal to the causal efficacy of phenomenal properties to defend the identity theory, rejecting multiple realizability.[4] To most, however, it seems too implausible to hold that only creatures with brains like ours get to be conscious.[5] The most popular move is to defend realization physicalism by rejecting (3). Some appeal to counterfactuals to show that the overdetermination involved in a phenomenal property and its physical realizer both causing the

[1] See Crane (1995) for the idea that the problem of mental causation should be treated as a mutually inconsistent set of plausible claims about the mind. I focus on phenomenal properties, but the problem generalises to other mental properties. See [Author1] for more on causal exclusion, especially in relation to functionalism.
[2] When I say that a property P causes an effect E, this should be taken to mean that P bestows upon E's cause C (which might be an event, state, or object) the *power* to cause E.
[3] Jackson & Pettit (1990).
[4] Lewis (1966); Kim (1992).
[5] Putnam (1967). Multiple realizability has been placed under significant pressure recently by Polger & Shapiro (2016). I lack the space to address their arguments here, but will briefly discuss their views in §2 and §4.

same effect is not of a problematic kind.[6] More recently, some have appealed to modal accounts of causation to argue that causation is stratified, so that mental and physical properties don't compete for efficacy. Mental causation of a token behaviour is implemented by lower-level causal relations between their token physical realizers, but because realized property-instances and their token realizers are indexed to different regions of modal space, the causal relations are strictly intra-level.[7] On such theories, despite the fact that the token physical realizers of a mental property-instance don't *cause* its effects, they nonetheless causally *determine* that those effects occur, by causing a token physical realizer thereof. This is clearly some kind of causal overdetermination, as the relevant effects are determined twice over, once by same-level mental causation, and again by lower-level physical causation plus non-casual realization.[8]

It's clear why realization physicalists are keen to defend overdetermination. According to Kim's widely accepted *causal inheritance principle*, realized properties inherit their causal powers from their physical realizers, and hence bestow a subset of those powers.[9] It follows that realized properties bestow powers to bring about effects that are caused (or causally determined, on layered accounts of causation) by their realizers, which is to say they are overdetermining causes of (some or all of) the effects of their realizers. Why embrace the inheritance principle? According to functionalist accounts of the mind, mental property M is individuated by causal role R, and to realize M is to fill R. Given that for a property to occupy a causal role is for it to bestow a certain set of causal powers, it's natural to say that if mental properties bestow powers themselves, those powers must be inherited from their realizers. Even accounts of realization that eschew functionalism respect some version of the causal

[6] Bennett (2003); Kallestrup (2006).
[7] Yablo (1992); List & Menzies (2010).
[8] Kim (1998) makes this point, and argues that the resulting overdetermination is no less problematic.
[9] Kim (1992).

inheritance principle.[10] On Wilson's view, for example, any physically acceptable theory of the mind-brain relationship ought to have as a consequence that the powers of a token mental property are a proper subset of the powers of its realizer on that occasion.[11]

My purpose in this paper is to suggest a theory of phenomenal properties that's consistent with all of (1)-(4). According to the solution I shall propose, the causal inheritance principle is false: phenomenal properties are multiply realizable properties that bestow causal powers their realizers do not, and so don't overdetermine their effects. This in turn requires that phenomenal properties are not entirely functional,[12] and that their realization doesn't consist solely in causal power bestowal. I shall argue that there are qualitative properties within the broadly physical domain, which are realized by more fundamental physical properties, yet don't inherit their causal powers from their realizers. If phenomenal properties are partially constituted by such properties, then (1)-(4) are consistent, if amended to distinguish two senses of 'physical'.

Take a *basic physical* property to be any simple, unrealized property that features in the laws of ideal completed physics, or any aggregate of such properties. Electric charge, mass, being an electron, and being two electrons 1m apart, are all basic physical on this definition. Now in addition to basic physical properties, there are also complex properties that can't be identified with aggregates of the properties of fundamental physics. Geometric shape is one example; neural synchrony, as we'll see, is another. Now if it turned out—as I maintain it does—that these latter properties made causal contributions that no other properties did, we wouldn't say they violated the causal closure of the physical. Physicalists don't interpret the 'physical' in

[10] Wilson (1999, 2011, 2015); Shoemaker (2001).
[11] This point is argued at length in Wilson (2015). See [Author2] for full discussion of causal inheritance as a condition on realization.
[12] On the theory suggested in §4, phenomenal properties are complex properties that are *partially* functional, and derive their causal novelty from their non-functional parts.

their closure principles to mean *basic* physical. Rather, they think that 'physical' means *broadly* physical, and refers to anything that is either basic physical, or stands in some appropriate relation to basic physical entities.[13] Here's a revised statement of the exclusion problem taking into account the distinction between basic and broad physicality:

1'. Phenomenal properties cause broadly physical effects
2'. The broadly physical domain is causally closed
3'. The effects of mental properties are not generally overdetermined
4'. Phenomenal properties are not identical to basic physical properties

I'll argue in what follows that phenomenal properties are broadly physical, multiply realizable qualities, which bestow novel causal powers in relation to their basic physical realizers.

2. Dependent Powerful Qualities: Realization without Inheritance

The powerful qualities ontology, developed by Heil and Martin, concerns basic physical properties.[14] The idea is that such properties are not pure powers as in Bird's dispositional essentialism,[15] nor are they intrinsically inert categorical properties as in Armstrong's Humean ontology.[16] Rather, they are both qualitative and powerful. Heil and Martin propose that the qualitative and the powerful are not aspects of fundamental properties, but identical. I find it difficult to make sense of this claim. As Jacobs understands them, powerful qualities are causally self-contained "thick quiddities": non-mental qualitative natures that are truthmakers of causal counterfactuals solely in virtue of being the natures they are.[17] In a similar vein, Smith argues for non-recombinatorial quidditism, according to which basic physical properties are qualitative natures that have certain causal roles solely in virtue of being the qualities they are,

[13] Crook & Gillett (2001).
[14] Heil (2003); Martin (2007).
[15] Bird (2007).
[16] Armstrong (1983).
[17] Jacobs (2011).

so that their existence at a world is sufficient for them to occupy those roles there.[18] However, once we say that powerful qualities are natures that ground their own powerfulness, it seems to follow that the qualities must be *prior* to the causal powers they bestow, which doesn't seem consistent with the identity claim.

The claim that powerful qualities ground their own causal roles is consistent with such qualities having partially causal essences, since a causal role that a property has in virtue of its essence may also be essential to it.[19] Even so, this way of treating powerful qualities looks like a dual-aspect theory, with the causal aspect grounded in the qualitative. For that reason, it seems unlikely that Heil or Martin would embrace it, but it's the treatment I shall adopt here. I'm not persuaded that basic physical properties are powerful qualities so understood, but I do think that some broadly physical properties are. In previous work, I have argued that molecular geometry is a dependent quality with an irreducible causal role. I shall briefly rehearse the arguments here in order to illustrate how realization without inheritance is possible.[20]

Water molecules have what we might call a basic physical structure—they consist of two hydrogen atoms covalently bonded to a single oxygen atom, standing in certain spatial relations to each other, having certain values of quantitative properties like mass and charge. But they also have a *geometric* structure: the two covalent O-H bonds lie at an angle of approximately 104.5° to each other, and this property is partially responsible for a range of characteristic behaviours, such as the disposition of water molecules to align in an electric field, and hydrogen bonding. We can define the geometric structure of the H_2O molecule as follows: the property

[18] Smith (2016).
[19] And of course it may not. See [Author3] for more on this issue.
[20] I present a simplified version of the argument here; see [Author2] for the details.

of being composed of proper parts *x*, *y* and *z*, arranged such that straight lines drawn between *x* and *z* and *y* and *z* subtend an angle of 104.5°. Fig.1 shows the structure of the H_2O molecule:[21]

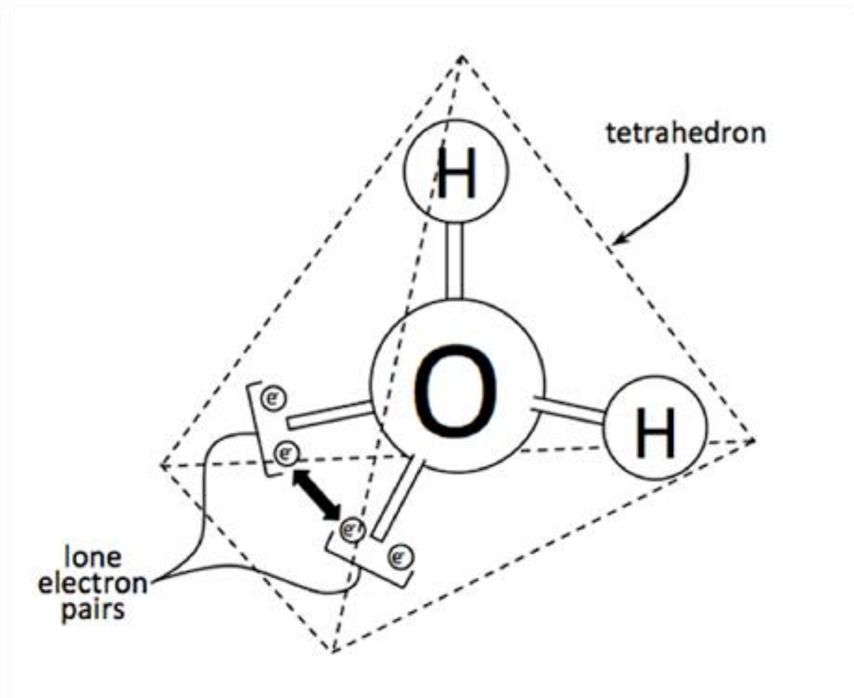

Figure 1 – Basic Physical vs. Geometric Structure of Water

The spatial relations between the atoms are causally determined, *inter alia*, by mutual repulsion between the two lone electron pairs and the two pairs of shared electrons in the covalent bonds, which form the vertices of a tetrahedron, as in the above diagram. Notice the difference between the basic physical and geometric structures of the molecule. The former is a structural compound of the specific spatial relations between the three atoms and electrons, the intrinsic properties of the various particles, and so forth. The latter, however, is a property common to both H_2O molecules and (for instance) macroscopic models thereof, whose basic physical structure is obviously very different. Having the H_2O molecule's basic physical structure is one way to have its geometric structure, but there are many others—the basic physical structure of

[21] Image courtesy of Wikipedia. File: Tetrahedral Structure of Water.png. (2017, February 20). https://commons.wikimedia.org/w/index.php?title=File:Tetrahedral_Structure_of_Water.png&oldid=234459610

H_2O *realizes* its geometric structure, and the same geometric structure is realized in a laboratory model by three plastic spheres standing in appropriate spatial relations. We need an account of realization that covers such cases, and I adopt the following:

> Property instance(s) $P_1,...,P_n$ realize a property-instance $F_\varphi(x)$ iff (i) x or its proper parts possess $P_1,...P_n$ in some combination; and (ii) x meets the specification φ definitive of F_φ in virtue of (i), but not *vice-versa*.[22]

I assume that realized dependent properties in general have definitions that explain their dependence on their realizers, and refer to the formulae that define them as *specifications*. The above account allows for functional realization—let φ be '*x* withstands a certain force without deformation', and let F_φ be the corresponding degree of rigidity. In this case, $P_1,...P_n$ will be properties like electric charge, and they will realize F_φ by bestowing upon *x*'s proper parts the powers required for *x* to occupy the φ-role. However, because some realized properties, such as geometric structure, are *not* causally defined, their realizers don't realize them by bestowing a set of causal powers. To have a tripartite bent geometry like that of the H_2O molecule, the three proper parts (arguably) must have some intrinsic properties or other, in order that they be capable of standing in spatial relations. In a molecular model, both the intrinsic properties of the relata *and* the spatial relations in which they stand are different to those found in H_2O molecules. What is common to the two structures is their abstract *geometric* structure.[23]

Is this kind of variation enough for "true" multiple realization? Polger and Shapiro argue that multiple realization is more demanding, hence less common, than is typically supposed.[24] Mere

[22] This account owes much to Gillett's (2003).

[23] A referee objects that molecular geometry is realized by "relative spatial relations", which are common to both H_2O molecules and molecular models. I am not sure what 'relative spatial relations' means, but it seems to me that it is most naturally taken to refer to the directions of the relata in relation to each other, in which case it is simply another way of referring to geometric structure. That is common to both cases, but its realizers—the relata themselves and the specific distance relations between them—are not.

[24] Polger & Shapiro (2016).

variations in properties like composition and size, such as between aluminium and steel waiter's corkscrews, won't suffice. For true multiple realization, there must be a difference in the way the defining function of a corkscrew is implemented. Variations in composition and size are lower-level variations in corkscrews of the same kind, *waiter's corkscrew*, not variations in the realization of the higher-level kind *corkscrew*.[25] It's not entirely clear to me what Polger and Shapiro would say about the realization of geometric properties, or other spatial and temporal patterns, not least because their account is explicitly restricted to the realization of functionally individuated kinds. More importantly, as we'll see in §4, whether or not the theory developed here theory entails true multiple realizability, it allows (*inter alia*) for conscious artificial intelligence, which is exactly the sort of thing whose apparent possibility typically motivates philosophers to reject the type identity theory and embrace functionalism.

It's one thing to argue that geometric structure is a dependent quality, but why suppose it to be powerful, and why suppose that its powerfulness stems from its nature, rather than from its realizers as in the case of functional properties? The answer to both questions is that there are certain causal facts concerning H_2O molecules that can't be explained without appealing to their geometric structure. Note that I'm not making the weak and relatively uncontroversial claim that geometric structure is needed in order to provide a certain *kind* of causal explanation of the relevant facts—a particularly simple or elegant explanation, perhaps. Rather, I have in mind the stronger claim that those facts can't be explained *at all*—at least not *fully*—without appealing to geometric structure. Let me explain why I think this. The H_2O molecule has a non-zero resultant dipole moment, which is a vector quantity resulting from the separation of (relatively) positive and negative charges, pointing by convention from positive to negative. In

[25] Gillett (2003) argues—and I agree—that because rigidity is itself multiply realizable, aluminium and steel waiter's corkscrews *do* count as different implementations of the same function.

each covalent O-H bond, the O atom attracts the shared electrons more strongly than the H atom, resulting in a bond with a dipole moment. The molecule as a whole is polar because there are components of these two vectors that point *in the same direction* and which therefore don't cancel each other out. Equivalently, we might say that the charge cloud is smeared out towards the O atom, leaving two relatively positive poles towards the H atoms, and a central relatively negative pole towards the O atom. The dipole moment of water is responsible for hydrogen bonding, which in turn is responsible for the fact that water is liquid at room temperature and 1atm pressure; and it's responsible for the disposition of H_2O molecules to align in an electric field, which is the principle upon which microwave ovens rely.

My central claim is that there's no way to explain why H_2O molecules have these powers without appealing to their geometric structure. Suppose we know that the charge cloud is drawn to the central O atom more strongly, resulting in a relatively negative O atom compared to the two H atoms. The spatial relations that realize the water molecule's geometric structure can be expressed in terms of the O-H bond length, B: each hydrogen atom is then a distance B from the O atom, and 1.58B from the other hydrogen atom. But why do *these* spatial relations explain the molecule's polarity, given what else we know? The simple answer is that they represent one way for the two H atoms to be located *in a certain direction* relative to the O atom, such that the two O-H bond dipoles don't fully cancel each other out.

We can deduce the dipole moment of H_2O from its basic physical structure, but only if we first use that information to deduce its *geometric* structure. That would be a very odd thing indeed, unless geometric structure were doing some genuinely novel causal work in relation to its basic physical realizers. It's hard to see how the role of geometric structure *could* be inherited from the basic physical properties that realize it on some occasion. It's only in combination that these properties suffice, in the circumstances, for a polar H_2O molecule, and the obvious explanation of this fact is that only in combination do they constitute a way for the two H atoms to be located in a certain direction relative to the O atom. In virtue of its basic physical structure, the H_2O molecule meets a geometric specification; and in virtue of meeting that specification, given its other properties, it has a net dipole moment. If there's causal inheritance here, it doesn't go from realizer to realized, but *the other way around*.

This move from explanatory relevance to causal novelty is likely to raise concerns.[26] There are plenty of properties that facilitate novel explanations without bestowing novel powers. As is familiar, explanations given in terms of functional properties provide for increased generality compared to explanations given in terms of the occupants of the relevant functional roles.[27] Functional explanations hold however the functional properties are realized, whereas the corresponding physical explanations apply only to physically similar causes and effects. Given that functional properties inherit their powers from their realizers, it can't be the case that a property having a novel causal-explanatory role suffices for its having novel causal powers. However, the present case is importantly different. When giving a functional explanation of some token effect, we know that we could (with loss of generality) explain that same effect solely in terms of the physical realizers of the relevant functional properties. In the present case, by contrast, it seems we cannot explain the dipole moment of H_2O without appealing to the geometric structure of the molecule. This in turn strongly suggests that geometric structure is as deep as the explanation goes: we can't do without it in the same way we can do without functional properties, and if we try, we lose much more than generality.

[26] I thank an anonymous referee for pressing me on this point.
[27] See [Author1] for full discussion.

Since geometric structure plays a unique role, alongside their basic physical properties, in determining how H_2O molecules behave, it's natural to say, following Shoemaker, that it bestows novel *conditional powers*.[28] Anything with the geometric structure of water will have a net dipole moment conditionally on being charged, having a central part that is more electronegative than the other two, and so forth. Philosophers tend to think that everything that happens in the physical world can be explained in wholly basic physical terms. Perhaps they think this because they think that all facts are ultimately grounded in basic physics, including the facts of causation, and conclude that all causal powers are bestowed by basic physical properties. Or perhaps they think this because they think that there's good evidence for the causal closure of the basic physical, which is violated if properties like geometric structure are causally novel. Let me address these worries in turn.

I reply to the grounding worry by accepting that the conditional powers of geometric structure are ultimately grounded in basic physics, but denying that this entails that those powers belong to basic physical properties. Basic physical properties ground dependent powerful qualities by realization, but realizing a powerful quality and bestowing its causal powers are not the same relation. Basic physical properties ultimately ground the causal powers of geometric structure, but *indirectly*, by means of two distinct grounding relations: (i) the realization relation between basic physical structure and geometric structure, and (ii) the direct bestowal relation between geometric structure and its conditional powers, which stem from its qualitative specification.[29] It may seem odd to attribute a novel causal role to the geometric structure of water when we can deduce this structure *a priori* from the spatial relations between the atoms. This, however,

[28] For something to have a power *simpliciter* is for it to be disposed to behave in a certain way, when appropriately related to certain stimuli. For something to have a conditional power is for it to be such that if it had certain other properties, it would have the relevant power *simpliciter*, where the other properties in question are not independently sufficient for this. See Shoemaker (2001), pp. 25-26.
[29] For details see [Author2]; for similar arguments see Gillett (2016a), pp. 221-3.

is a consequence of the fact that geometric structure is realized by basic physical structure in a fully transparent way, due to the explicit specification of the former. Functional properties are also *a priori* deducible from their realizers: given that $P_1,...P_n$ bestow upon *x*'s proper parts the powers needed for *x* to occupy the φ-role, it follows *a priori* from *x*'s proper parts having $P_1,...P_n$ that *x* has the functional property F_φ of being such as to occupy the φ-role. That F_φ is causally redundant, however, is a consequence of its being specified in terms of a causal role that's *occupied* by $P_1,...P_n$, not its mere *deducibility* from $P_1,...P_n$. Dependent qualities like geometric structure aren't functionally defined, and so don't inherit this principled limitation.

Second, the causal closure worry. The evidence physicalists typically take to support the causal closure of the physical involves successful causal explanations of physical phenomena without appealing to anything outside the *broadly* physical domain. Such explanations frequently involve non-basic, broadly physical properties, and there's no evidence that I know of that such properties can always be eliminated in favour of purely basic physical explanations—quite the contrary. There is a kind of downward causation involved here, however, that violates the causal closure of the *basic* physical domain. A water molecule's geometric structure is partially responsible for its dipole moment, and that in turn affects the motions of its basic physical proper parts. This may seem troubling, but it needn't be. Geometric structure exerts no special forces, so my view is consistent with the well-supported principle that all forces are generated by basic physical properties such as electric charge. But the ways in which such forces manifest depend, irreducibly, on the abstract spatial patterns formed by basic physical particulars.

Now it might be suggested at this point that it's incorrect to describe the geometric structure of H_2O as bestowing causal *powers*. Genuine causal powers are powers to exert forces, and the admission that geometric structure doesn't exert forces might seem tantamount to admitting that it doesn't do any causal work. We might say, for instance, that rather than bestowing causal powers, geometric structure is a condition on the *manifestation* of causal powers that are bestowed by the molecule's basic physical properties. This is similar to Shoemaker's account of emergence, according to which basic physical properties bestow *micro-latent* powers whose manifestation requires a certain kind of complex structure.[30] The difference is that Shoemaker denies that the effects of micro-latent powers are deducible from basic physical structure. In the present case, there's no failure of deducibility, because molecular geometry is deducible from basic physics, and the behaviour of the H_2O molecule can be fully explained by appealing to its basic physical properties and geometric structure together.[31]

Alternatively, we might say, following Gillett,[32] that basic physical properties only *bestow* some of their causal powers when they realize certain dependent broadly physical properties. For Gillett, dependent properties are causally individuated properties of complex systems, and realizing them consists in joint role-filling by the lower-level properties of the system's components. In cases Gillett refers to as *machresis*, a dependent property non-causally determines that its realizers bestow certain conditional powers, and is itself instantiated in virtue of those very powers, by joint role-filling. Because it makes a difference to the powers of its realizers, Gillett treats the dependent property as a joint cause of their effects—but the powers it manifests are fully composed by the powers of its realizers. This theory allows us to say that the geometric structure of H_2O determines that its basic physical realizers bestow certain causal powers they otherwise would not, such as the power to align in an electric field.[33]

[30] Shoemaker (2002).
[31] See [Author2] for more on this issue.
[32] Gillett (2016a,b).
[33] It isn't clear where properties like geometric structure fit into Gillett's ontology, and I don't attribute to him the claim that they are capable of machresis.

There's much more to be said about these issues, but I lack the space to say it here. For present purposes, I need only note that both of the above alternatives entail that properties like geometric structure do *something* that their realizers don't, differing only as to *what it is they do*. Any genuinely novel causal role for mental properties will solve the exclusion problem, whether it consist in empowering or conditioning. I say that dependent powerful qualities have their causal roles in virtue of being the properties they are, and this too is something upon which the above alternatives can agree. Suppose, as in Gillett's account, that the basic physical properties of H_2O bestow all its conditional powers, but that some of them are conditional on its geometric structure. Although in this case the conditioning role of geometric structure in H_2O is grounded in the powers its basic physical realizers bestow, it doesn't *inherit* that role from them, because they don't *occupy* it. It's geometric structure, not its realizers, that does the conditioning—and it does so by being the very property it is, because *that's* the property that water's basic physical properties somehow pick out as a condition on their causal powers. A machretic condition makes a difference to the powers of its realizers, and it doesn't inherit *that* role from them, even though it inherits its causal powers. Machretic conditions thus have a role they don't inherit from their realizers, and which makes a difference to the course of events. Few would worry about causal exclusion if it turned out mental properties had such a role.

3. Neural Synchrony as a Dependent Powerful Quality

Neurons *oscillate*. These oscillations can be rhythmic variations in the sub-threshold membrane potential, or rhythmic sequences of action potentials (*spike trains*). I shall focus on spike trains, but my central conclusions may also be applicable to sub-threshold oscillations. Sub-threshold membrane potential is around -70mV. Neurons below threshold pump Na^+ ions out of the cell and K^+ ions in across the cell membrane, resulting in a relatively high concentration of K^+ inside the cell body, and a relatively high concentration of Na^+ outside. An action potential is triggered when post-synaptic potentials combine to raise the membrane potential of the cell to around -55mV. Sodium channels first open in the cell membrane, resulting in a rapid diffusion of Na^+ ions into the cell body, which raises its electric potential (depolarization). This process is sequential, with local depolarizations triggering the opening of adjacent sodium channels. Potassium channels then open, resulting in a diffusion of K^+ ions out of the cell, returning it to just below resting potential (hyperpolarization). Ion pumps then restore the resting concentration of Na^+ and K^+ ions inside and outside the cell membrane.

Action potentials are thus rapid changes in the membrane potential of a neuron, which travel down the axon until they reach the axon terminal, where the change in local potential causes inhibitory or excitatory changes to the membrane potential of post-synaptic neurons via chemical or electrical synapses. Sensory neurons modulate their firing rate in response to stimuli in their receptive fields. For instance, retinal ganglia in the visual system receive input from large numbers of rod and cone cells (photoreceptors in the retinae) and can be either 'on-centre' or 'off-centre'. On-centre ganglia fire rapidly when, and only when, the rod and cone cells in their receptive fields are illuminated by a central light spot surrounded by a dark border, whereas off-centre cells do the opposite. These cells are used by the visual system to detect changes in contrast, and hence the edges of objects. The information in question is carried by oscillations: neurons communicate patterns of changes in their membrane potential, such as firing rates, to each other. Neuronal oscillations can also occur in entire populations. These oscillations are fluctuations in the average number of action potentials per unit time, and constitute oscillations in the network's electric potential, since each individual spike is a rapid rise and fall in the membrane potential of an individual neuron. Fig.2 displays a simulation of a neural population oscillating at 10Hz. The dots in the upper part represent action potentials, with the lower part representing the overall oscillation in local field potential that results from their rhythmic density fluctuations. Call this *local neural synchrony*:[34]

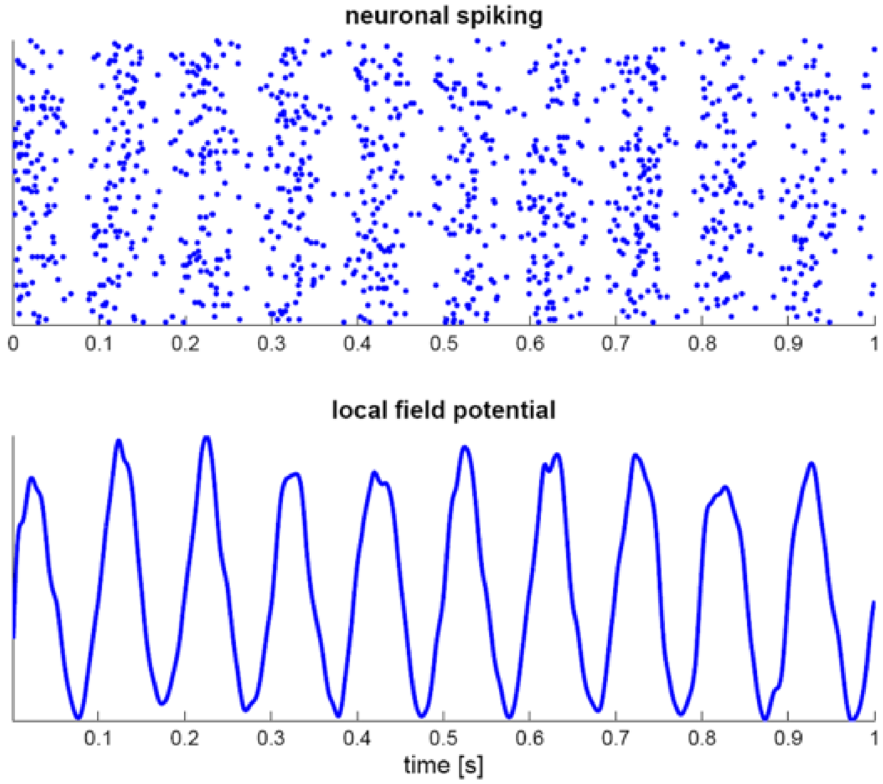

Figure 2 – Schematic Representation of Local Neural Synchrony

Within such an oscillatory pattern, further oscillations can be encoded. Suppose, for instance, that the density of spikes varies not only at 10Hz, but also at 5Hz, with odd numbered peaks having a particularly high spike density compared to evens. In this case, local field potential would still peak every 0.1s, but the peaks would vary rhythmically in height at a slower rhythm of 5Hz. Multiple rhythms can thus be encoded by neuronal oscillations. These rhythms can be communicated to other populations, and processed by them, resulting in distinct oscillatory patterns. There are many forms of neural synchrony, with the simplest involving oscillation at the same frequency and in phase. However, synchrony can also occur with non-zero phase lag,

[34] Image courtesy of Wikipedia. File:SimulationNeuralOscillations.png. (2011, September 7). https://commons.wikimedia.org/w/index.php?title=File:SimulationNeuralOscillations.png&oldid=59030332.

and between neural populations oscillating at different frequencies.[35] It occurs not only locally, within particular cortical areas such as the visual cortex, but also *non-locally*, between distinct areas responsible for different cognitive and sensory tasks, such as the visual and parietal cortices.[36] Neural synchrony has been shown in multiple studies to correlate with conscious experience and attention.[37]

When the subject visually attends to a stimulus, neurons in the visual cortex have been shown to oscillate synchronously in the gamma range (20-100Hz, but 40Hz is typical).[38] In the pyramidal-interneuron network gamma (PING) model, local *gamma coherence* is sustained by interaction between excitatory neurons and inhibitory interneurons. Roughly, an excitatory input that activates a network in the visual cortex also triggers local inhibitory interneurons, which render neurons less susceptible to excitatory inputs until inhibition decays, after which the network is again sensitive to excitatory input. This results in a rhythmic variation in the excitability of the network.[39] The *frequency* of oscillation is determined by factors such as the intrinsic tendency of the relevant neurons to oscillate in their sub-threshold membrane potential, the delay before inhibitory neurons are triggered, and the length of the period of inhibition.[40] The mechanisms by which these oscillations become synchronized in neural networks are no doubt highly complex and varied, but this kind of behaviour is typical of coupled oscillators. Think of the well-known phenomenon, discovered by Huygens in 1665, in which two pendulum clocks hanging on a beam become spontaneously synchronized so that they swing at the same frequency and 180° out of phase. Similarly, interactions between

[35] I focus on same-frequency oscillations for now. I return to the issue of phase lag below, and briefly discuss cross-frequency coupling in §5.
[36] Engel et. al. (2001); Fries (2005).
[37] I return to this in §4.
[38] Gray & Singer (1989); Fries et. al. (2001).
[39] See Fries (2015), pp. 221-2; Singer (2013) pp. 6-7, for summaries; and Tiesinga & Sejnowski (2009) for details.
[40] Tiesinga & Sejnowski (2009); Wang (2010).

coupled oscillators in a neural network give rise to spontaneous local neural synchrony. Locally coherent populations can then in principle entrain other populations to the same rhythm by similar means, giving rise to *long-range* neural synchrony between them.[41]

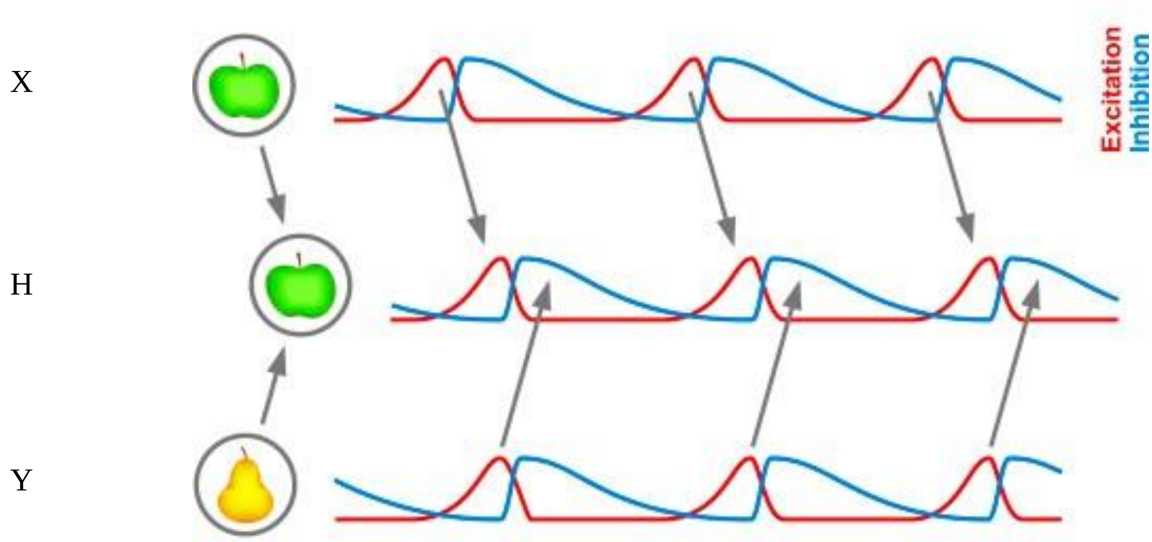

Figure 3 – Communication Through Coherence

Neural synchrony is now widely held[42] to be a central mechanism by which neurons are bound into efficient functional assemblies. Refer to fig.3.[43] Suppose two populations X and Y of neurons in the visual cortex represent two distinct objects, an apple and a pear respectively, by means of two distinct spatial patterns of locally synchronous gamma band oscillations. Suppose further that X and Y project to a higher visual area H, but that X and H are gamma coherent, whereas Y and H are not. Apple representations from X will arrive at H during excitability peaks, whereas pear representations from Y won't activate H because they arrive during the inhibitory part of the gamma cycle. Furthermore, local gamma coherence within X means that

[41] In some models, long-range gamma coherence is established by a chain of the same inhibitory interneuron networks that give rise to local gamma coherence in the PING model; in others, it is mediated by slower rhythmic oscillations that act as pacemakers. For details, see Buzsáki & Wang (2012). To get a sense of the intricacies of synchronization between coupled oscillators in general, see Arenas et. al. (2008).
[42] Following Fries' influential (2005) paper.
[43] Reproduced from Fries (2015).

its neurons tend to fire together, maximising its excitatory input to H. Hence the representation of the apple in X is efficiently communicated to H, while the representation of the pear in Y is not. Communication between gamma coherent populations is thus both efficient and selective. This is the core of the communication-through-coherence hypothesis (CTC).[44]

There's a lot more to be said about CTC than I can say here. What's important for present purposes is to note the importance of temporal patterns in the explanation of how neural synchrony facilitates effective communication. For CTC to work, the frequency and relative phase of the oscillating populations must be such that signals from one population always arrive during the peak excitability phase of the other, before the inhibitory part of the cycle begins. How two or more populations achieve this will depend not only on their oscillatory frequencies (which need not be the same), but also on their anatomical distance apart in the brain, and on the nature of the coupling between them, all of which will vary considerably between populations. Neural synchrony, I suggest, is a dependent powerful quality: it isn't basic physical, but it plays an irreducible role in determining the course of neural events.

The present case is a temporal analogue of the H_2O case, in which (roughly) the frequencies of oscillation correspond to the spatial relations between atoms, and the relative phase corresponds to the angle between the bonds. Interactions between the proper parts of the H_2O molecule cause it to have a geometric structure that's partially responsible for the way water behaves. Likewise, interactions between oscillating populations cause them to oscillate synchronously, and synchrony itself then plays a crucial role in explaining the selectivity and efficiency of communication between them. In the H_2O case, there's no way to fully explain the various causal properties of the molecule without appealing to the spatial angle between

[44] Fries (2005, 2015).

the two O-H bonds. Similarly, in CTC there's no way to fully explain efficient communication between neural populations without appealing to the *phase* angle between their oscillations. Just as geometric structure bestows various conditional powers on the H_2O molecule, so too neural synchrony bestows upon neural oscillators the power to communicate in a particularly efficient way, conditionally on the instantiation of other properties. One such property is being such that their sensitivity to post-synaptic inputs is correlated with phase. Otherwise, efficiency of communication wouldn't depend on when in the phase of oscillation excitatory signals arrived. Another such property is being suitably coupled. Coincidental synchronization, for instance between neural populations in the brains of two different individuals, obviously doesn't suffice for the power *simpliciter* to communicate efficiently and selectively.

An immediate worry about CTC is that because neural synchrony is causally *dependent* on communication, it can't causally *explain* communication. CTC requires coupling relations by means of which the coherent populations interact, and which cause and sustain synchronous oscillations. It's the fact that X and H are engaged in an on-going process of communication that explains their synchrony. X activates H, which involves communication of X's activity pattern to H, and thereby triggers the local inhibition that results in (i) H not responding to input from Y, and (after inhibition decays) (ii) H being ready to receive the next burst of excitatory input from X. But how can synchrony explain communication between X and H, if it also depends on it? First, note that there's no reason why we can't appeal to the fact that X and H are synchronous to explain why Y is unable to communicate with H, so there's no circularity in the claim that X and H achieve selectivity via synchrony. Second, synchrony can explain distinct forward-looking cognitive properties of X and H. For example, synchrony explains how X can *maintain* its apple representation in H over an extended period, and why

H is particularly sensitive to *changes* in the contents of X's apple representation, for instance due to the apple being in motion, or a fly landing on it, during that time.[45]

Many of the properties that interest neuroscientists are multiply realizable, broadly physical properties. Some, such as excitability, are functional; others, such as neural synchrony, are not. Within the brain, neural synchrony can occur at different frequencies, between various populations with different coupling relations, and with different phase lags corresponding to their anatomical separation. Neural synchrony is plausibly multiply realized in the brain, but this doesn't show that *consciousness* is multiply realizable. That depends on how synchrony and consciousness are related. In §4 I suggest a theory according to which phenomenal properties are representational contents realized by synchronous oscillators. I further suggest that consciousness is constitutively tied not to *neural* synchrony, but to synchronous oscillation *per se*, which gives significant scope for variation in the physical natures of conscious systems.

4. Neural Synchrony and Consciousness

Neuroscientists have a range of techniques at their disposal to measure synchrony, and track the ways in which it correlates with various types of conscious experience. They can use micro-electrodes to measure local field potentials (electric potential in the space between neurons) at specific areas of the brain and see how they change during cognitive tasks that require conscious attention; or they can use non-invasive techniques such as MEG and EEG to measure evoked magnetic and electric fields, respectively, during such tasks. Mathematical analysis of the data gathered in these ways can reveal increases in synchronous oscillation during

[45] Fries (2015), p. 224, suggests a pulsatile coding system in which spatial activity patterns carry representational content and are refreshed in the receiving population at the frequency of oscillation.

conscious cognitive tasks. I can't offer a comprehensive summary here, but will give some relevant examples. The papers cited refer to many further studies for the interested reader.

In binocular rivalry, two incongruent images are presented, one to each eye. The result is bistable perceptions that switch from one image to the other. Subjects report when their conscious state switches, and their brains are monitored to see which neural changes correlate with changes in conscious state. In one study,[46] a red vertical grating was presented to one eye, and a blue horizontal grating to the other. The two gratings were made to flicker at different frequencies, with the frequency then "tagging" neural populations with the given stimulus in their receptive field, which oscillate at the same frequency.[47] Subjects were asked to activate a switch with the left index finger whenever the red grating was dominant, and another switch with the right index finger whenever the blue grating was dominant. An array of MEG sensors was then used to record modulation of synchrony in stimulus-evoked magnetic fields at the flicker frequency corresponding to each stimulus during binocular rivalry. The authors found that the conscious percept was strongly correlated with increased synchrony at the stimulus frequency, not only within the visual cortices, but also between distinct cortical areas.

As the authors put it, "coherence between distant MEG sensors reflects the level of synchronization between different brain regions."[48] Stimulus-evoked magnetic fields at a given flicker-frequency are the result of summed action potentials in neurons with the relevant stimulus in their receptive fields, so such results indicate that conscious percept correlates with increased neural synchrony in binocular rivalry. Other binocular rivalry studies come to a

[46] Srinivasan et. al. (1999).
[47] The frequency tags were in the range 7-12Hz. Such flicker-induced oscillations are not those that the brain uses to represent the gratings themselves, but a different rhythm occurring in the same network. As explained in §3, neural networks can oscillate at several frequencies at once.
[48] Op. Cit. p. 5446.

similar conclusion, providing evidence that conscious percept correlates much more strongly with increased synchronization between areas representing the stimuli than it does with firing rates. In a study using implanted electrodes to measure local field potentials during binocular rivalry in awake striabismic[49] cats, early visual neurons representing the stimulus shown to the dominant eye were found to increase in their gamma coherence, while those representing the stimulus shown to the suppressed eye decreased theirs. Interestingly, no correlation between conscious percept and firing rates of the representing neurons was found in these areas, which seem to represent the stimuli shown to both the dominant and suppressed eyes continuously.[50]

Other studies indicate a correlation between consciousness and long-range gamma coherence between distinct cortical areas. In one such study,[51] human subjects were asked to identify whether or not a word briefly presented for 33ms was the same as a word presented 533ms later. The first word was preceded and followed by a masking stimulus of duration 67ms, whose luminance was varied to determine whether or not the first word was reported as consciously perceived by test subjects. The authors measured evoked potential from a wide range of brain areas using EEG, and found that in both the conscious and non-conscious cases, there were increased local gamma oscillations. In the non-conscious case, the word was still represented, as confirmed by priming effects related to the meaning of the unperceived word. However, exclusively in cases where the first word was reported as perceived, there was a short period of gamma coherence measured by electrodes at widely separated cortical areas, including occipital, parietal and frontal areas. The occipital lobe is home to the early visual cortices and is responsible for detecting features like edges, colours, and motion. The parietal lobe has many

[49] In this condition one eye is dominant and the other suppressed when different stimuli of equal contrast are presented to each eye, and this fact can be used in a similar way to binocular rivalry studies, but rather than relying on reports, we simply assume that the dominant eye image is perceived during rivalry.
[50] Fries et. al. (1997). Details of further such studies can be found in Engel et. al. (1999); for a recent and wide-ranging review, see Engel & Fries (2016).
[51] Melloni et. al (2007).

functions, including object interaction, touch, and egocentric spatial location. And the frontal lobes are associated with perceptual memory formation, judgement, planning, motor control, and language. The fact that conscious awareness correlates with long-range neural synchrony among such diverse cognitive systems is clearly a significant result.

Finally, there is evidence that synchrony is correlated with top-down attentional focus, which we know from our experience to be correlated with a higher degree of conscious awareness of the attended stimulus. For instance, one study found that attention to a visual stimulus increased the local gamma coherence of visual cortex V4 neurons representing that stimulus.[52] In another, that involved monkeys switching attention between a visual task and three distinct tactile tasks of varying difficulty, local synchrony in somatosensory cortex SII was found to increase when performing the tactile task, and to increase proportionally depending on the difficulty of the task.[53] In general, just as low-level visual representations can entrain higher level visual areas to achieve greater perceptual salience compared to other representations, so, conversely, top-down attentional processes can entrain lower-level populations to raise the salience levels of their representations, for instance in visual search.[54]

These and other experiments show a correlation between neural synchrony and consciousness, but correlation doesn't imply constitution, and many extant theories explain why neural synchrony is corelated with consciousness without positing a constitutive connection.[55] On such theories, the fact that neural synchrony is a dependent powerful quality won't help us with

[52] Fries et. al. (2001).
[53] Roy et. al. (2007).
[54] See Engel et. al. (2001) for a detailed review of the role of neural synchrony in top-down attentional processing. They suggest that visual search might work by top-down induced synchrony in the sub-threshold membrane potential of visual populations, so that when their preferred stimulus is detected, its representation is automatically synchronous with the higher cognitive systems guiding the search.
[55] I won't summarise them here, but see Engel & Fries (2016) for a survey of several such theories.

the causal exclusion problem, because it doesn't help explain how phenomenal properties could be both multiply realizable and causally novel. In what follows, I'll offer an alternative theory according to which synchronous oscillation is partially constitutive of conscious experience. My primary aim isn't to defend the theory in question, but to highlight one way in which neural synchrony might be constitutive of consciousness, thereby showing that the exclusion problem can be solved without giving up on multiple realizability. The specific proposal is probably too simplistic to be correct, but it isn't the only way to develop a theory that posits a constitutive link between oscillatory patterns and consciousness. Theories that posit more sophisticated connections will solve the exclusion problem in distinctive ways, depending on the kind of causal novelty the relevant patterns possess, and how they are related to consciousness.

Following Chalmers, take a *pure* representational property to be the property of representing an object, property, event, state of affairs,…, and an *impure* representational property to be the property of representing an object in a certain specific manner.[56] Pure representationalism is then the thesis that phenomenal properties are identical to pure representational properties, and impure representationalism the thesis that they are identical to impure representational properties. As Chalmers notes, pure representationalism is implausible. Since phenomenal properties are essentially conscious, pure representationalism forces us to either deny that there could be unconscious intentional content, or posit a special kind of content that's always conscious. It's more plausible that phenomenal character is intentional content whose object is represented in a phenomenal manner. Within impure representationalism, one can be a

[56] Chalmers (2004). Chalmers speaks of pure and impure representational properties as representations *of content* rather than states of affairs. I don't think this is a particularly perspicuous way of putting things, and prefer to think of intentional states as representing objects (properties, events…), not contents.

reductionist about content, or the phenomenal manner of representation, or both, or neither.[57] The view I propose here is a doubly reductionist form of representationalism.

Consider first Tye's PANIC theory of consciousness, according to which phenomenal properties are Poised, Abstract, Nonconceptual, Intentional Contents.[58] Tye endorses a causal covariation account of content, according to which a state S of *x* represents that $p =_{df}$ If optimal conditions obtain, S is tokened in *x* if and only if *p* and because *p*.[59] For Tye, it's the property of being *poised* that is the phenomenal manner of representation; the other conditions specify the *type* of contents that are conscious when their objects are so represented. Representations are poised, for Tye, in that they "stand ready and available to make a direct impact on beliefs and/or desires", and as Tye himself notes, this is a functionalist criterion.[60] On Tye's theory: (i) states have their intentional contents in virtue of causal covariation under optimal conditions, (ii) states with abstract, non-conceptual contents are phenomenally conscious when they also possess a certain functional role in relation to beliefs and desires. Tye's theory clearly faces the causal exclusion problem, because both contents and the phenomenal manner of representation are understood in broadly functionalist terms.

There's an obvious similarity between Tye's representationalism and the *global workspace theory* (GWT), which treats phenomenal consciousness in terms of the global availability of content.[61] GWT is a version of representationalism in which the phenomenal manner of representation is analysed in terms of global availability. Being poised is a form of availability

[57] One can also be a Russellian or a Fregean representationalist, depending on whether one identifies phenomenal character with phenomenally representing an object *simpliciter*, or with representing it under a given mode of presentation. For present purposes, I'm neutral between these alternatives. See Chalmers (2004).
[58] Tye (1995).
[59] Op. cit. p. 101.
[60] Tye (2000), p. 62.
[61] GWT was first proposed in Baars (1988).

that falls somewhat short of being global, as it involves only the disposition to directly cause belief and desire states, but there are clear affinities. According to GWT:

> [A]t any given time, many modular cerebral networks are active in parallel and process information in an unconscious manner. An information becomes conscious, however, if the neural population that represents it is mobilized by top-down attentional amplification into a brain-scale state of coherent activity that involves many neurons distributed throughout the brain. The long-distance connectivity of these 'workspace neurons' can, when they are active for a minimal duration, make the information available to a variety of processes including perceptual categorization, long-term memorization, evaluation, and intentional action. We postulate that this global availability of information through the workspace is what we subjectively experience as a conscious state.[62]

Some notes are in order. First, the global workspace isn't a special place in the brain—a Cartesian Theatre, to borrow Dennett's term—where contents become conscious. Rather, the idea is that (for instance) visual representations become conscious *right where they are* when made available to neural systems that categorize the object, assess its significance, store information about it, and use that information to guide behaviour.[63] Second, there needn't be anything special about the other cognitive systems availability to which makes a certain content conscious. What's important is that the content is widely available, rather than the type of representation or processing that occurs in the relevant populations. On this version of the GWT, the fact that two systems make a given content available *to each other* may render that content to some degree conscious, without either doing any special consciousness-engendering processing.[64] Third, it's the global *availability* of content that matters to consciousness, not how widely tokened that content actually is. As Dennett's puts it, the phenomenal manner of representation in GWT is analysed in terms of *clout*, not *fame*.[65]

[62] Dehaene & Naccache (2001), p. 1.
[63] This is the central message of Dennett (1991).
[64] This suggestion is made in Dennett (2001), p. 224, and I'll return to it later.
[65] Op. cit., pp. 224-5.

If CTC gets the cognitive role of neural synchrony right, then synchrony is among the *realizers* of the global workspace envisaged by many as the basis of consciousness.[66] However, the phenomenal manner of representation is a second-order functional property in GWT: to represent an object phenomenally is for its representation to be globally available, never mind how this availability is realized. Thus, GWT is a theory that explains the correlation between neural synchrony and consciousness without positing a constitutive link. In what follows, I'll suggest identifying the phenomenal manner of representation with neural synchrony itself, rather than with functional properties it realizes. I assume that content is multiply realizable, and allow for the sake of argument that content properties inherit their powers from their realizers. Multiple realization holds not only on causal covariation theories such as Tye's, but also on teleosemantic theories, according to which an organism's states represent not those features or states they actually covary with, but those they were *selected* to covary with in the evolutionary history of the organism;[67] and on isomorphism theories such as Cummins's, according to which states represent anything with which they are structurally isomorphic.[68] Before proceeding, let's consider the relationship between content and neuronal oscillations.

Assume that Tye's theory is true, on which to represent the content that *p* is to be a state S that is tokened iff *p*, and because *p* (under ideal conditions). Focus on the visual system. Visual cortex neurons modulate their firing rates according to the presence or absence of certain stimuli, responding differentially to upstream feature-detection neurons in their receptive fields, until we get to the sensory receptors at the periphery. For instance, neurons in V1 are tuned to simple features like edges, V2 is tuned to properties like colour and shape, V5 to motion, and so forth. Neurons in these areas oscillate more rapidly when their preferred stimuli

[66] The suggestion that synchrony realizes the global workspace is made in Newman & Baars (1993).
[67] Millikan (1984); Papineau (1984).
[68] Cummins (1996).

are detected, and represent them by oscillating. Because *p* is a propositional content, S will a complex state composed of oscillating neurons in a range of distinct specialized populations. On Tye's theory, basic feature-detecting neurons represent the features detected in virtue of their disposition to be present iff and because the features are present, with complex propositional representations built from feature representations by means of a compositional semantics.[69] Hence, neuronal oscillations *realize* content on Tye's account.

I lack the space for a full treatment, but the same will be true on any informational semantics, including teleological accounts, since neuronal oscillations, *qua* feature detectors, are the carriers of the relevant information, and plausibly selected to covary with the features they detect. And the same is also true on Cummins's account, although the notion of realization employed will be different, to take account of the fact that content isn't a functional or dispositional property.[70] On Cummins's account, structures carry information about states of affairs, with elements of the structure inheriting their contents from the content of the whole, whose contents in turn stem from isomorphism. The structures in question are the same patterns of neuronal oscillation that realize content properties on indicator semantics, except that now they get their content from their structure rather than their dispositional or teleological properties. Since structure is a relatively abstract property, the concrete neuronal oscillations that instantiate it in us are best seen as its realizers.[71] On each of these theories, content is multiply realizable, and plausibly realized in us by neuronal oscillations.

[69] It's independently plausible that the visual system uses a combinatorial semantics, which in turn gives rise to the feature binding problem. See Treisman (1999) for an introduction, and the other papers in the same volume for further discussion.
[70] The account given in §2 allows for the realization of abstract structure.
[71] Content so understood may also be causally novel, if the arguments of §2 are correct. A structural representation may have certain causal powers in virtue of the very same abstract structure that endows it with representational content, and not in virtue of the realizers of that structure. I lack the space to develop this proposal here, however, and shall not assume any causal novelty on the part of content.

The proposal I want to consider here is that to represent an object phenomenally is for it: (i) to be represented by means of oscillations in potential, and (ii) to be represented by perceptual and (a sufficient number of) higher cognitive systems, suitably coupled to each other, and by means of that coupling synchronous. While the accounts of content sketched above place no specific constraints on its realizers, the theory just mooted places constraints on the kinds of systems that could represent things *phenomenally*—but not, in my view, constraints that preclude multiple realizability. Why tie phenomenal representation to synchrony *per se*, rather than *neural* synchrony? As noted earlier, neurons communicate patterns of changes in their membrane potentials to each other via action potentials. Information about how threshold is reached in a given neuron, and the mechanism of the action potential, is screened off by the fact that post-synaptic neurons respond only to changes in their membrane potential. On the assumption that consciousness is grounded in synchronous representation, it would be very odd if it somehow also depended on the precise implementation of synchrony in the brain, when only the electrical changes themselves are communicated between neurons.

Assume for illustration a bottom-up (or 'feedforward') theory of perceptual representation, according to which (for instance) visual contents are initially computed by and tokened in the visual cortices, then passed to higher cognitive areas. Given the above theory, an apple is then represented phenomenally when the visual cortex neural populations that represent its perceptible features are engaged in long-range synchronous oscillations with higher cognitive networks, such as those responsible for: classifying it as an apple, and as food; representing it as desirable; planning future actions in the light of the available information; and so on. As we've seen, synchronous oscillation in the brain is driven by communication between neural populations, so on the current proposal, contra-Dennett, the phenomenal manner of representation requires *fame*, and not just *clout*. Phenomenal representation so understood is a process in which perceptual and cognitive populations achieve synchrony by means of the communication of representations, and sustain efficient and selective communication by means of synchrony.

The present theory doesn't suffer from the causal exclusion problem, because synchronous oscillation—hence the phenomenal manner of representation—is a causally novel, broadly physical property, and not a purely functional property like global availability. Why represent a given object phenomenally? Because so representing it enables the content in question to be preferentially communicated, on an on-going basis, to cognitive systems that can use it for a range of important purposes, and because it affords those same systems quick access to any changes in the properties of the object represented. If I am correct that these powers are not inherited, then phenomenal consciousness has a genuinely novel causal role that seems a good fit for our pre-theoretical conception of what consciousness is *for*. There's clearly a lot more to be said about these issues; let me gesture at some of it here.

First, the current model allows phenomenal consciousness to differ both in degree and in kind depending, respectively, on how many and on which higher cognitive systems are synchronous with the populations responsible for perceptual discrimination. The model could be extended to cover cases such as consciously thinking about the meaning of a word, because the neural populations that represent meanings can also in principle exhibit the kind of synchrony involved in the phenomenal manner of perceptual representation. As in Dennett's version of GWT, we can allow that populations represent phenomenally in virtue of being synchronous *with each other*, rather than thinking of specialized representational systems supplying the content to higher areas, which supply the consciousness. This in turn allows for many different degrees and kinds of consciousness depending on which systems are involved. Relatedly, the theory also allows for more realistic versions of representationalism, in which both bottom-up and top-down processes determine the contents of perception. We might say, for example, that contents *distributed* between perceptual and higher cognitive systems are conscious when those systems are suitably coupled and thereby synchronous with each other.

Second, there are several cases in which we see neural synchrony without associated states of consciousness, for example so-called hypersynchrony during epileptic seizures,[72] and more importantly for present purposes, there's evidence of increased long-range gamma coherence during general anaesthesia.[73] The first case shows that synchronous oscillation *simpliciter* isn't sufficient for consciousness, but I don't claim that it is. The second case shows that gamma coherence is also not sufficient for consciousness, but again, I don't claim that it is. What I have suggested is that long-range synchrony between neural circuits tokening perceptual contents, and those responsible for higher cognitive functions, is sufficient for the content in question to be conscious. That doesn't commit me to finding consciousness wherever we find synchrony; to represent phenomenally is to represent the world in a way that constitutively involves neural synchrony, which allows for non-conscious synchronization.

Third, the proposal developed here is consistent with the popular idea that the brain uses temporal coding for feature binding. According to an influential model,[74] top-down attentional processes bring about gamma coherence in certain feature detecting networks, which binds them into object representations. Only when so unified are these representations capable of activating working memory, which is where they become conscious. I reject the second part of this theory, but I needn't reject the first. As we saw in §3, neural populations can oscillate at a

[72] Rao & Lowenstein (2015).
[73] Murphy et. al. (2011). I thank an anonymous referee for drawing my attention to this.
[74] Crick and Koch (1990); see also Engel et. al. (1999).

many different frequencies at once, which allows for multiple cognitive roles for temporal coding. I can also see no reason in principle why the brain could not use local gamma coherence within the visual cortices for visual feature binding, and long-range gamma coherence to represent the bound features phenomenally.

Fourth, given the kind of impure representationalism suggested here, phenomenal representation is a complex process involving both functional and qualitative properties. Content is plausibly a functional property, and so is the property of being coupled, which enables and sustains long-range neural synchrony.[75] While it matters to me that all features of phenomenal representation are multiply realizable, it isn't problematic if some of them fail to be causally novel. The causal novelty of phenomenal states doesn't stem from functional properties such as their contents, but from lower-level temporal patterns between token representations and higher cognitive systems, and such patterns, I have argued, are dependent powerful qualities for which the causal inheritance principle is false.

Fifth, and finally, let's reconsider multiple realizability. Block famously argues that functionalism is too *liberal* when it comes to the attribution of consciousness.[76] Anything with the right causal structure will be conscious on functionalist theories, and that doesn't seem right—as Block suggests, it seems we could set up a large enough group of people with suitable signalling devices to instantiate the causal structure of human psychology, but without there being anything that it's like to *be* that group. Reductive representationalism of the kind envisaged here places broadly physical constraints on conscious systems, so instantiating a psychological causal structure isn't sufficient for consciousness. It's also not necessary. If two

[75] I thank an anonymous referee for pressing me on the nature of these couplings.
[76] Block (1974). Searle (1980) famously raises the same concern for the classical computational theory of mind, but we can safely view this as a version of functionalism for present purposes.

oscillating populations can represent phenomenally, to some degree, by means of synchrony with each other, then consciousness is much easier to construct than human psychology, and quite likely a commonplace in the natural world. In order to produce conscious systems, all Mother Nature has to do is evolve systems that represent by means of oscillations in their electric potential, and couple them to each other in appropriate ways. Local and global synchrony will then arise spontaneously, so She gets consciousness for free.

Block also objects to *psychofunctionalist* theories, which tie consciousness closely to human brain structure, on the grounds that they are too chauvinistic.[77] This kind of objection doesn't bite here, however, because the physical constraints imposed aren't *too* strict. In principle, we should be able to construct silicon networks that represent objects and their properties by means of oscillations in potential, and couple them to further oscillatory networks that process these representations in various ways. On the present proposal, phenomenal consciousness ought to emerge spontaneously in such a system. Polger and Shapiro[78] endorse a 'modest identity theory' according to which some variations in brain properties between conscious agents—for instance, variations in neuronal mass or ion concentration, and other variations that don't meet their criteria for multiple realization—are consistent with the mind-brain identity theory. On my view, phenomenal properties aren't identified with neural properties at all, but with oscillatory patterns that can in principle be instantiated by systems without biological neurons. Even if this isn't true multiple realization, it's something near enough.

[77] Block (1974).
[78] Polger & Shapiro (2016).

5. Conclusion

The simple coherence-based version of representationalism presented above is no more than a sketch of a theory of consciousness, but the idea that oscillatory patterns are constitutively linked to consciousness can be extended to more sophisticated accounts.[79] In predictive coding models of consciousness, for instance, conscious perception is treated in terms of the construction of theoretical models to predict primary data from sensory receptors, such as a map of the wavelength composition and luminance of light incident on the retinae. These predictive models are radically underdetermined by the data, and much 'guesswork' is needed to construct them.[80] The predictive structure is hierarchical, with each level constructing progressively more sophisticated models of the data at the level below. Data from sensory systems flow upwards in the hierarchy, giving rise to models that attempt to predict those data, where prediction is essentially suppression of the lower-level signals in question. Input from lower sensory areas that isn't predicted continues to flow upwards as an error signal, with the network as a whole attempting to minimise error, thereby generating better guesses as to what the world would have to be like in order for the sensory signals to be as they are.

It has been suggested that gamma oscillations are involved in the bottom-up transmission of sensory data, while the top-down predictive signals are carried by slower beta oscillations.[81] On this account, the gamma coherence we observe to be correlated with conscious experience corresponds to the flow of unpredicted data upwards in the hierarchy, which constrains the construction of predictive models. If we identify conscious perception with the entire process of model construction, then we might still find a constitutive role for gamma coherence in

[79] It also doesn't require a commitment to representationalism about phenomenal character, although I'll continue to assume representationalism here

[80] See Friston & Stephen (2007) for details..

[81] Bastos et. al. (2012).

representing phenomenally, assuming gamma coherent oscillations encoding lower-level representations are involved in their upward transmission.[82] Alternatively, we might see the process of model construction as largely unconscious, with conscious perception occurring as its output once prediction error has been minimised. In that case we might say that the phenomenal manner of representation constitutively involves cross-frequency coupling relations between oscillations carrying top-down and bottom-up representations, with properties like the relative phase and frequency of these oscillations playing a crucial role in the suppression of sensory data by predictive models. On such accounts, we will need to look beyond CTC to fully describe the causal novelty of consciousness, but the general point stands: if phenomenal properties are identified with spatial and temporal patterns of oscillation, then in principle they can be both multiply realizable and causally novel.

The attraction of functionalist accounts of mental properties is multiple realizability, and it's natural to suppose that once we move away from functionalism and identify those properties with what occupies a given role rather than the role itself, we lose multiple realizability. Conversely, if we want multiple realization, we're stuck with functionalism and the causal exclusion problem. Not so, in my view. The basic physical domain is not causally closed, because there are at least some causally novel broadly physical properties. But broadly physical properties—such as geometric structure, neural synchrony, and in general the kind of spatial and temporal patterns with which the special sciences are concerned—can be realized in many different ways. To go functional, then, is to go too far: there are multiply realizable properties

[82] It's worth noting that some who endorse predictive coding models of perception eschew mental representation altogether, for instance Hutto (2018). See Kiefer & Hohwy (2018) for a diametrically opposed view, according to which mental representation, construed in terms of causal-structural isomorphism, arises naturally within Bayesian predictive coding systems.

within the broadly physical domain that are also causally novel, and it's these properties we should look to in order to solve the causal exclusion problem.[83]

[83] Based primarily on research funded by the Fundação para a Ciência e a Tecnologia (IF/01736/2014), and in part on earlier work funded by a British Academy Postdoctoral Fellowship. I am very grateful to Peter Dayan for discussion of downward causation in neuroscience, and for introducing me to CTC, but do not attribute any of the views presented here to him. For helpful discussions and comments, I am grateful to David Chalmers, Carl Gillett, Matteo Grasso, Jim Hopkins, William Jaworski, Robert Koons, David Papineau, and two anonymous referees.

References

Arenas, A., Diaz-Guilera, A., Kurths, J., Moreno, Y. and Zhou, C. (2008). 'Synchronization in Complex Networks', *Physics Reports* 469, pp. 93-153.

Armstrong, D. (1983). *What is a Law of Nature?* Cambridge University Press.

Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

Bastos, A., Usrey, M., Adams, R., Mangun, G., Fries, P., Friston, K. (2012). 'Canonical Microcircuits for Predictive Coding', *Neuron* 76, pp. 695-711.

Bennett, K. (2003). 'Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract It', *Noûs* 37, pp. 471-497.

Bird, A. (2007). *Nature's Metaphysics*. Oxford University Press.

Block, N. (1978). 'Troubles with Functionalism', *Minnesota Studies in the Philosophy of Science* , pp. 261-325.

Buzsáki, G. & Wang, X.-J. (2012). 'Mechanisms of Gamma Oscillations', *Annu. Rev. Neurosci.* 35, pp. 203-25.

Chalmers, D. (2004). 'The Representational Character of Experience', in B. Leiter (ed.) *The Future for Philosophy* (pp. 135-181). Oxford University Press.

Crane, T. (1995). 'Mental Causation', *Proceedings of the Aristotelian Society* supp. vol. 69, pp. 211-236.

Crick, F. and Koch, C. (1990). 'Towards a Neurobiological Theory of Consciousness', *Seminars in the Neurosciences* 2, pp. 263-275.

Crook, S. and Gillett, C. (2001). 'Why Physics Alone Cannot Define the 'Physical', *Canadian Journal of Philosophy*, 31(3), pp. 333-359.

Cummins, R. (1996). *Representations, Targets, and Attitudes*. Cambridge: MIT Press.

Dehaene, S. & Naccache, L. (2001). 'Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework', *Cognition*, 79, pp. 1-37.

Dennett, D. (1991). *Consciousness Explained*. Little, Brown & Company.

Dennett, D. (2001). 'Are we Explaining Consciousness Yet?', *Cognition*, 79, pp. 221-237.

Engel, A., Fries, P., Konig, P., Brecht, M. and Singer, W. (1999). 'Temporal Binding, Binocular Rivalry and Consciousness', *Consciousness and Cognition* 8, pp. 128-151.

Engel, A., Fries, P. and Singer, W. (2001). 'Dynamic Predictions: Oscillations and Synchrony in Top down Processing', *Nature Reviews Neuroscience* 2, pp. 704-715.

Engel, A. and Fries, P. (2016). 'Neuronal Oscillations, Coherence and Consciousness', in S. Laureys, O. Gosseries and G. Tononi (eds.) *The Neurology of Consciousness* 2nd ed. (pp. 49-60). Academic Press.

Fries, P., Roelfsema, P., Engel, A., König, P. and Singer, W. (1997). 'Synchronization of Oscillatory Responses in Visual Cortex Correlates with Perception in Interocular Rivalry', *Proc. Natl. Acad. Sci. USA* 94, pp. 12699–12704.

Fries, P., Reynolds, J., Rorie, A. and Desimone, R. (2001). 'Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention', *Science* 291, pp. 1560-1563.

Fries, P. (2005). 'A Mechanism for Cognitive Dynamics: Neuronal Communication through Neuronal Coherence', *TRENDS in Cognitive Sciences* 9(10), pp. 474-480.

Fries, P. (2015). 'Rhythms for Cognition: Communication through Coherence', *Neuron* 88, p-p. 220-235.

Friston, K. & Stephan, K. (2007). 'Free Energy and the Brain', *Synthese* 159, pp. 417-458.

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

Gillett, C. (2016a). *Reduction and Emergence in Science and Philosophy*. Cambridge University Press.

Gillett, C. (2016b). 'Scientific Emergentism and its Move Beyond (Direct) Downward Causation', in Paoletti, M. & Orilia, F. (eds.) *Philosophical and Scientific Perspectives on Downward Causation*, Routledge, pp. 242-262.

Gray, C. and Singer, W. (1989). 'Stimulus-specific Neuronal Oscillations in Orientation Columns of Cat Visual Cortex', *Proc. Natl. Acad. Sci* 86, pp. 1698-1702.

Heil, J. (2003). *From an Ontological Point of View*. Oxford University Press.

Jackson, F. & Pettit, P. (1990). 'Program Explanation: a General Perspective', *Analysis*, 50(2), pp. 107-117.

Jacobs, J. (2011). 'Powerful Qualities, Not Pure Powers', *The Monist* 94, pp. 81-102.

Kallestrup, J. (2006). 'The Causal Exclusion Argument', *Philosophical Studies* 131, pp. 459-485.

Kiefer, A. & Hohwy, J. (2018). 'Content and Misrepresentation in Hierarchical Generative Models', *Synthese* 195, pp. 2387-2415.

Kim, J. (1992). 'Multiple Realization and the Metaphysics of Reduction', *Philosophy and Phenomenological Research*, 52(1), pp. 1-26.

Kim, J. (1998). *Mind in a Physical World*. MIT Press.

Lewis, D. (1966). 'An Argument for the Identity Theory', *Journal of Philosophy* 63, pp. 17-25.

List, C. and Menzies, P. (2010). 'The Causal Autonomy of the Special Sciences', in C. Macdonald and G. Macdonald (eds.), *Emergence in Mind* (pp. 108-128). Oxford University Press.

Martin, C. B. (2007). *The Mind in Nature*. Oxford University Press.

Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W. and Rodriguez, E. (2007). 'Synchronization of Neural Activity across Cortical Areas Correlates with Conscious Perception', *The Journal of Neuroscience* 27, pp. 2858-2865.

Millikan, R. (1984). *Language, Thought and Other Biological Categories*. MIT Press.

Murphy, M., Bruno, M., Riedner, B., Boveroux, P., Noirhomme, Q., Landsness, E., Brichant, J., Phillips, C., Massimini, M., Laureys, S., Tononi, G., & Boly, M. (2011). 'Propofol Anaesthesia and Sleep: a High-density EEG Study', *Sleep* 34, pp. 283-291.

Newman, J. and Baars, B. (1993). 'A Neural Attentional Model for Access to Consciousness: A Global Workspace Perspective', *Concepts in Neuroscience* 4, pp. 255-290.

Papineau, D. (1984). 'Representation and Explanation', *Philosophy of Science* 51, pp. 550-72.

Polger, T. & Shapiro, S. (2016). The Multiple Realization Book. Oxford University Press.

Putnam, H. (1967). 'The Nature of Mental States', in W. H. Capitan & D. D. Merrill (eds.), *Art, Mind, and Religion* (Pittsburgh University Press), pp. 37-48.

Rao, V. & Lowenstein, D. (2015). 'Epilepsy', *Current Biology* 25, R742-R746.

Roy, A., Stelnmetz, P., Hsiao, S., Johnson, K., & Niebur, E. (2007). 'Synchrony: A Neural Correlate of Somatosensory Attention', *J. Neurophysiol.* 98, pp. 1645-1661.

Searle, J. (1980). 'Minds, Brains and Programs', *Behavioural and Brain Sciences* 3, pp. 417-24.

Shoemaker, S. (2001). 'Realization and Mental Causation', *Proceedings of the Twentieth Word Congress of Philosophy* 9, pp. 23-33.

Shoemaker, S. (2002). 'Kim on Emergence', *Philosophical Studies* 108, pp. 55-63.

Singer, W. (2013). 'The Neuronal Correlate of Consciousness: Unity in Time rather than Space?', in A. Battro, S. Dehaene & W. Singer (eds.) *Neurosciences and the Human Person: New Perspectives on Human Activities*, Pontifical Academy of Sciences *Scripta Varia* 121, pp. 51-67.

Smith, D. C. (2016). 'Quid Quidditism Est?', *Erkenntnis* 82, pp. 237-57.

Srinivasan, R., Russell, P., Edelman, G. and Tononi, G. (1999). 'Increased Synchronization of Neuromagnetic Responses during Conscious Perception', *The Journal of Neuroscience* 19, pp. 5435-5448.

Tiesinga, P. and Sejnowski, T. (2009). 'Cortical Enlightenment: are Attentional Gamma Oscillations Driven by ING or PING?', *Neuron* 63, pp. 727-732.

Treisman, A. (1999). 'Solutions to the Binding Problem: Progress through Controversy and Convergence', *Neuron* 24, pp. 105-110.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press.

Tye, M. (2000). *Consciousness, Colour and Content*. MIT Press.

Wang, X.-J. (2010). 'Neurophysiological and Computational Principles of Cortical Rhythms in Cognition', *Physiol. Rev*. 90, pp. 1195-1268.

Wilson, J. (1999). 'How Superduper does a Physicalist Supervenience Need to Be?', *Philosophical Quarterly*, 50(194), pp. 33-52.

Wilson, J. (2011). 'Non-reductive Realization and the Powers-based Subset Strategy', *The Monist*, 94(1), pp. 121-154.

Wilson, J. (2015). 'Metaphysical Emergence: Weak and Strong', in T. Bigaj and C. Wüthrich (eds.), *Metaphysics in Contemporary Physics* (pp. 345-402). Poznan Studies in the Philosophy of the Sciences and the Humanities.

Yablo, S. (1992). 'Mental Causation', Philosophical Review, 101(2), pp. 245-280.

Yates, D. (2012). 'Functionalism and the Metaphysics of Causal Exclusion', *Philosophers' Imprint* 12(13), pp. 1-25.

Yates, D. (2016). 'Demystifying Emergence', *Ergo* (3) 31, pp. 809-841.

Yates, D. (2017). 'Inverse Functionalism and the Individuation of Powers', *Synthese*. DOI: 10.1007/s11229-017-1417-9.