MANUSCRITO
International Journal of Philosophy

Article submitted to the similarity system

turnitin

Original Articles

CAUSAL-PATTERN THEORIES OF CONSCIOUSNESS: A CHALLENGE AND A META-CAUSAL RESPONSE

JOHN BARNDEN
https://orcid.org/0000-0001-8062-2288
Emeritus Professor of AI
University of Birmingham
School of Computer Science
Birmingham
UK
jabarnden@btinternet.com

Article info
CDD: 126
Received: 21.03.2024; Revised: 07.11.2024; Accepted: 07.11.2024
https://doi.org/10.1590/0100-6045.2024.V47N1.JB

Keywords
Phenomenal Consciousness
Meta-Causation
Process Theories of Consciousness
Causal Efficacy of Consciousness
Causal-Pattern Theories of Consciousness

Abstract: This article presents a challenge concerning the causal efficacy of causal processes, distinct from the much-discussed causal-exclusion problem. The new challenge is to consciousness theories that require conscious processes to involve causation patterned in some specific way. This broad, diverse class includes prominent theories such as the Integrated Information Theory, Global Workspace theories and a type of Higher-Order Thought theory. The challenge arises because the causal pattern is not itself required for the effects the processes have on the organism's other aspects. Hence, the processes' property of being conscious is dispensable in accounting for those effects. The theories are challenged to show how this does not constitute an operational problem for individual organisms or a problem as regards explaining the evolution of consciousness. The paper explains how the challenge can be met by the radical move of introducing meta-causation: causation that acts on causation instances themselves. This allows the instances of causation in conscious processes, as entities in their own right, to be causes of effects elsewhere. The paper also summarizes the author's previously published motivation for proposing meta-causation, as the basis of consciousness itself. The present paper further supports this view.

Acknowledgements

I am grateful for thoughtful and constructive criticism from the anonymous reviewers, leading to many improvements to the paper.

1. Introduction

A large and long-standing issue within the study of [phenomenal] consciousness[1] is whether consciousness is

[1] By "consciousness" I will always mean phenomenal consciousness. An entity is phenomenally conscious at some moment in its existence if there is "something that it is like" to be that entity in its current state—it feels like something, to the entity, to be currently existing. By contrast there is (I assume) nothing it is like *to* the paving stones on one's patio to be currently existing. Casting phenomenal consciousness as what-it-is-like-ness or what-it-feels-like-ness is fairly standard in the study of consciousness, at

epiphenomenal or not (Robinson 2023). This largely concerns the question of whether people's or other organisms' conscious mental states or processes are causally efficacious upon the physical world. Do they causally affect physical circumstances beyond (i.e., outside or subsequent to) that state or process, whether these circumstances are within the organism itself or within the external environment? Or are conscious states or processes just "pointless" side-effects or concomitants of the physical world? (That would be the epiphenomenal case.) Henceforth, by the brief term *causal efficacy* of a conscious process I will mean the sort of efficacy on other, physical, circumstances as just described, unless otherwise noted.

In the present paper I address what seems to be a largely neglected aspect of the causal-efficacy issue, concerning the role that causation lying *within* (and coming into) a conscious process plays in the process's causal efficacy as just defined. A background assumption here is that consciousness is indeed a feature of processes—it's a matter of activity, rather than of activity-less states (whether instantaneous or extended in time). The activity is driven by the causation within the process, combined with incoming causation. The activity might, for instance, be the changing activation levels (and other aspects of state) of the neurons in some particular neuron network within some brain, with the internal causation arising by virtue of the signals transmitted between neurons, though the paper has a much broader scope than this case.

The paper makes the working assumption that it is indeed desirable to have a consciousness theory that equips conscious processes with causal efficacy. On this basis, the main thrust of this paper is, first, to point out a challenge to a certain important class of consciousness theories, as regards whether and how they can adequately provide the causal efficacy in such a way as to fit with the idea that

---

least since Nagel (1974).

consciousness arose as a result of evolution. The challenge is just a challenge, and not a knock-down argument against the theories, but I believe the relevant theorists would do well to try explicitly to meet it. The second half of the paper's main thrust is to propose a particular way to meet the challenge.

I call the challenge the *Physical-Level Causal Dispensability* (PLCD) challenge, to be outlined later in this Introduction. The reason it is not widely addressed[2] is perhaps that attention has been taken up with a somewhat analogous but distinctly different problem, the so-called causal-exclusion problem for mental states, to be discussed briefly in section 6.1.

My proposed way of meeting the challenge is the radical move of giving *meta-causation* a crucial role in consciousness theory. Meta-causation, also known as "higher-order" causation or "iterated" causation (not to be confused with chained causation), is where cases of causation can in themselves be causes or can be causally affected. An intuitive example of the notion, at the level of everyday personal interactions, could be a mother being caused to be angry by her son John causing her daughter Mary to cry. The latter *causing* of a change in Mary by some action of John's is here itself a cause, so the causing-mother-to-be-angry is meta-causation. Meta-causation is little discussed even in mainstream causation theory (as attested also by Kovacs 2021); and, at least in the physical form that I advocate, has to my knowledge not been given a role in consciousness theory prior to the work leading up to this paper (Barnden 2014, 2020, 2022).[3] However, I will discuss some related

[2] But it was briefly mentioned by Ignacio Cea (2023) in a talk as a problem facing the Integrated Informtaion Theory.

[3] In Barnden (2014) and (2020) I preferred terms other than "causation" and "meta-causation" to put forward my approach, in an effort to avoid preconceptions about causation. I respectively used "[meta-]running" and "[meta-]dynamism". My current use of "[meta-]causation" is only a terminological change and does not

notions, including downwards causation, in other philosophical work in the Discussion section (section 6). Note that I present the PLCD challenge as being of interest in its own right, irrespective of any merits or otherwise of the meta-causal response to it. Conversely, I submit that meta-causation is of interest separately from the PLCD challenge. In this light, I briefly discuss my prior motivation for introducing meta-causation into consciousness theory, informing my prior work just cited. This motivation did not involve the PLCD challenge. Meeting that challenge can be seen as a new, additional benefit of the previously developed meta-causal approach.

In the following comments in this Introduction I will motivate the working assumption of causal efficacy that was mentioned above, then give the flavour of the PLCD challenge, and then point out briefly that a meta-causal approach could meet it. I postpone further description of meta-causation itself until section 4.

There are reasonable grounds for supposing that consciousness theories should indeed seek to equip conscious processes with causal efficacy. Putting aside our intuitions that our conscious states do cause us to do things, these grounds inlcude, on the one hand, some empirical evidence that conscious processes in people and some animals do have causal efficacy, and, on the other hand, the arguable desirability of having a theory of consciousness that answers an evolutionary question: the problem of explaining how consciousness could have arisen, through processes of evolution, from an entirely non-conscious early, physical state of the world. Such an explanation is aided if conscious processes have causal efficacy, especially if that efficacy is not readily achievable by otherwise-comparable non-conscious processes (more on this later). If we had a theory that satisfyingly answered the evolutionary question, then the

reflect a change of content. My notion of causation is similar to the notion of "oomph" (Demarest 2017, Kutach 2014, Schaffer 2016).

scientific and philosophical understanding of the world would be more complete, and I simply assume here that this would be a desirable outcome.

There has been extensive discussion in the biological and related literature on claimed or possible functions of conscious processing in people and existing or past animals (e.g., Birch 2022, Brown 2022, Cleeremans & Tallon-Baudry 2022, Kolodny, Moyal & Edelman 2021, Crump & Birch 2022, Jablonka & Ginsburg 2022, Ludwig 2022, Marchetti 2022, Newen & Montemayor 2023, Niikawa *et al* 2022). Discussed functions may be functions of conscious processing in general or of particular aspects of it such as pain. Generally speaking, it seems fair to take talk of a function of a conscious process to be at least implicitly about its causal efficacy in our sense. For instance, much of the discussion of pain is about what it causes the organism to do.

An important tranche of the discussion of functions is about consciousness in non-human animals, and one important development in this area is Birch's (2022) proposal that we should take, as a marker of conscious- ness in an animal, the presence of a cluster of cognitive abilities that appear to be facilitated by conscious processing. By facilitated Birch means at least partially caused. Birch provisionally proposes a particular cluster of abilities, but what those are, or what the finally agreed cluster might be, is not central to this paper, which is about how causal efficacy in general is achieved. Birch's proposal is based on his Facilitation Hypothesis, which states that "Phenomenally conscious perception of a stimulus facilitates, relative to unconscious perception, a cluster of cognitive abilities in relation to that stimulus." This concentrates on (conscious) perception, whereas of course consciousness is a much broader matter, but that need not detain us, as conscious perception is probably the aspect of consciousness that the literature focusses on most. Birch states that the Hypothesis is consistent with a wide range of theories of consciousness. This accords with my own impression, although the matter is not clear-cut. As Newen & Montemayor (2023) indicate, theories are often unclear or just silent about what the evolutionarily relevant functions of consciousness within the

broader organism are meant to be.

Dung (2023) elaborates on the use of Birch's approach for the purpose of tests for consciousness in animals and machines. Significantly for us, he says "the entire project of finding tests of consciousness does pre- suppose that consciousness makes some functional difference, i.e., has systematic causal effects." And this project is indeed a popular one. We should also note the work of Ben-Haim *et al* (2021), suggesting that an organism may behave oppositely in the case of conscious awareness of a stimulus versus unconscious perception of it.[4] Ben-Haim *et al.* casts this as a "double dissociation" between conscious and unconscious perception.

Discussions of the functions of conscious processing implicitly or explicitly raise the question of why and how consciousness arose in the course of evolution on our planet. I will simply assume here that it did so, or at least that relatively advanced forms of it in human beings and some other animals did so. I also assume that it arose as a result of normal evolutionary processes rather than through some as-yet-undiscovered evolutionary mechanism. Some of the above-referenced literature on functions of conscious processing explicitly considers evolution and some does not. But in either case the default assumption, or at least working hypothesis, appears to be that if conscious processing (all of it, or advanced forms of it) did arise through evolution then it was because of some adaptive role it had/has, i.e. a role beneficial to the species in question. However, we should remain aware that it could instead have been a useless or even harmful side-effect of a beneficial development. If consciousness did have a beneficial role that drove its evolution, then it is usually assumed, at least tacitly, that the role is a matter of conscious processes causing individuals of the species to behave in certain beneficial ways in certain circumstances. Thus, evolutionary considerations support, without definitely implying, the causal efficacy of conscious

[4] However, a caution: that article does not explicitly mention causation at all.

processing. (However, see Robinson 2023 for further discussion.)

These comments leave room for the possibility arising in panpsychism that consciousness of some primitive sort has always existed in basic forms of physical matter. (See Brüntrup & Jaskolla 2016 for various forms of panpsychism.) Such a theory leaves it open that more complex, animal-level consciousness still arose through evolution, so that in effect evolution solved a form of the famous "combination problem" that panpsychism faces of how to combine basic conscious elements into a more complex conscious whole. The considerations of this paper are therefore relevant to this panpsychic possibility as well as to the possibility that consciousness evolved out of an entirely consciousness-free world. Somewhat similarly, the paper leaves room for biopsychism (see Lamme 2022 for discussion), the idea that all forms of life are conscious. It is still presumably the case that the more advanced forms of consciousness had to evolve out of lower forms.[5]

Turning now to the PLCD challenge, I need first to clarify what class of consciousness theories it is a challenge to. It is a challenge to theories that implicitly or explicitly claim that one necessary feature of a conscious process is that the causation within it and pushing it along, perhaps together with causal influences into the process from outside, satisfies some particular condition. I call the combination of a process's internal causation and causation coming into to it from outside itself the process's *governing causation*. I assume that the theory requires the same condition on a process' governing causation across all types of entity that might be considered—it's not that there's one condition for people and another for bees and another for robots, for example. I will say that such a theory requires a conscious process's governing causation to be "patterned" in a certain way---or, more briefly, that the process include a certain type of causal pattern. I will therefore call the theories

[5] But, to clarify, my own theory of consciousness is neither panpsychic nor biopsychic.

causal-pattern theories. This is not to imply that the theories themselves focus sharply on the required causal patterning: they may be much more concerned with other matters. One theory that does explicitly highlight causal patterning is, famously, the Integrated Information Theory (IIT).[6] But I also include, under causal-pattern theories, the Global Workspace approaches and one sort of Higher-Order Thought theory ("actualist" ones), even though causation is not nearly so explicitly salient or central there. I will explain why the PLCD challenge arises for these three types of theory in section 3. But I would suggest that many other theories are also vulnerable to it. Anyway, the theories just mentioned are, scientifically and philosophically, highly prominent, widely-advocated theories of consciousness, rightly or wrongly.[7]

The PLCD challenge itself is somewhat complex and nuanced, and is explained fully in section 3. It can conveniently be presented as having two aspects, though they are highly interlinked. These aspects are in essence as follows:

---

[6] It is the prime example of "causal-structure theories" in the terminology of Doerig *et al* (2019). My "causal-pattern" terminology has roughly the same intent, but I disagree with Doerig *et al* on which theories are included, as they explicitly exclude the Global Workspace approach.

[7] Causal-pattern theories constrain the causation *within* conscious processes, and this is a very different idea from causal functionalism, which postulates that the nature of a mental state is determined by its causal effects on, and causal influenceability by, *other* states. (See, e.g., Levin 2023. I am counting a conscious process as a mental state.) Such a theory automatically builds in the claim that conscious processes have causal efficacy, and escapes the PLCD challenge as long as it does not postulate conditions on causal patterns within conscious processes. However, I do not wish to rely on causal functionalism as a way for a causal-pattern theory to meet the PLCD challenge, on grounds beyond the scope of this paper.

- *Core of the challenge: Consciousness and consciousness-compliant causation is causally dispensable:-* In the theories in question, a conscious process's causal efficacy can be accounted for without bringing in the process's compliance with the special causal-pattern condition that the theory imposes. This is simply because what is directly responsible for the process's causal efficacy is something other than that causation, namely one or more states reached by the process. The process's governing causation is therefore "dispensable" in an account of the process's causal efficacy. As a further consequence, the process's being-conscious is itself dispensable in an account of the process's causal efficacy.[8]
- *Specific implications: There are difficulties concerning the needed structure of individual organisms and the evolution of types of organism:-* The causal dispensability makes it more difficult to see how consciousness could have arisen by selection in evolution. A particular point here is that there may well be non-consciousness-compliant alternatives to the causal patterns required by causal pattern theories. The states directly responsible for the causal efficacy of a conscious process could well have arisen through non-conscious processing. These alternatives (a) might exist in types of organism other than a conscious type that is under consideration, and/or (b) might exist alongside the conscious processing in a given type of organism. In either (a) or (b) we get evolutionary difficulties. Also, aside from evolutionary questions, in (b) we get special difficulties concerning how the processing within a given organism needs to be organized.

The core aspect taken by itself is of philosophical importance, and challenges a theory to be clear about what its stance on the issue is: e.g,, that the theory is admitted to fall short of desired causal efficacy for consciousness and therefore needs further development, or that the theory is unconcerned about the issue. The needed clarity may need

[8] The problem here is exacerbated by the nature of current mainstream philosophy of causation, as I explain in section 4.3.

to be accompanied by clarity on which of the many theories of causation in philosophy, or some new one, is being appealed to by the theory (as consciousness theories are typically unclear on this point).

However, this paper is more concerned with the specific implications mentioned in the second bullet. They constitute a considerable enlargement of the challenge beyond the core aspect.

The plan of the rest of the paper is as follows.

Section 2 mentions an additional distinction between types of theory (identity theories and non-identity theories), cutting across the distinction between causal-pattern theories and others. The distinction is needed in order to highlight an important complication in how the claims of the paper need to be couched. Section 3 presents the PLCD challenge in detail, and briefly exemplifies the form that the core aspect of challenge takes in IIT, GWTs and HOT theories (see above). Section 4 explains this paper's meta-causal response to the challenge. It also looks briefly at the prospects for getting existing relevant theories of causation to include meta-causation.

The sections so far constitute the main contributions of the paper. But Section 5, for completeness, proceeds to outline my different, prior motivation for introducing meta-causation (Barnden 2020, 2022; with preliminary considerations in Barnden 2014). There, I claim, on the basis of considerations separate from the PLCD challenge, that consciousness is actually itself based on a suitable arrangement of meta-causation. I claim that all conscious processes include a particular type of pattern of meta-causation, and meta-causation is not just a way in which conscious processes interact with circumstances beyond themselves. In short, the claim is that we need a *meta*-causal-pattern theory of consciousness. (I have been developing a particular theory of this sort, called MDyn—see Barnden 2020, and also Barnden 2022.) The prior motivation does not

depend on the validity of the PLCD challenge or on the meta-causal approach to it; and there is no dependence in the other direction either. However, the fact that the prior meta-causal approach to consciousness meets the PLCD challenge can be seen as a new benefit of that approach.

Section 6 is a discussion section, and starts with discussing a partial analogy with, but crucial differences from, the much debated "causal exclusion" problem for some theories. It also discusses some further matters, including notions related to meta-causation and possible challenges to the idea that the world includes any meta-causation at all, let alone meta-causation within conscious processes. Section 7 concludes.

Note that, while the main explicit focus of the paper is on biological organisms on our planet, it has some implications for the study of conscious entities of any other kind that might be considered, be they conscious AI systems if there are or will be any, natural non-biological conscious physical systems if there are any, conscious alien beings if there are any, or some other type of conscious entity (even non-physical ones). For brevity, I will use the term "organism" to cover all cases unless otherwise indicated.

2. Identity Theories and Non-Identity Theories

This paper considers "conscious processes" in the sense of physical processes that have the property of being conscious. However, it is neutral as to whether the property of being conscious is itself considered to be a physical property or not, and, if it is a physical property, whether it is physical in the sense of being identical to a condition expressible in neutral physical terminology—i.e., physical terminology that does not itself rely on consciousness (e.g., current terminology in physics)— or is instead physical in some special, extended sense, on a physical level ontologically above what is normally regarded as physical. However, the paper does assume a tight relationship between being-conscious and some neutral physical property, in that it assumes that an entity is conscious *if and*

only if it satisfies the condition. Thus, the paper allows, at one extreme, mind/brain [type-]identity theories, taken to be theories where being-conscious is identical to being in a condition expressible in the language of neurophysiology; and towards the other extreme it allows being-conscious to be strongly emergent from the physical (as in non-reductive physicalism, under the meaning of that term in Robb, Heil & Gibb 2023), and, more distantly still, it allows consciousness to lie in some entirely non-physical realm, as long as there is sufficient coordination of some sort with the physical realm to secure the above if-and-only-if.

I will use the term "identity theory" to cover theories where being-conscious is identical to *some* sort of neutral physical condition, irrespective of whether this is a biological condition or not (as in mind/brain identity theories). For instance, it might be a condition expressible in terms of basic physics and requiring specific sorts of field etc. in specific configurations. Or it might be a very general condition that many specific types of physical system, perhaps involving different sorts of physical matter or field, could satisfy. I will use "non-identity theory" to mean remaining theories that secure the above if-and-only-if. Of course, identity theories trivially secure it.

The identity/non-identity distinction cuts across the distinction between causal-pattern theories and other theories. This paper is concerned with causal-pattern theories irrespective of whether they are identity or non-identity theories. In either the identity case or the non-identity case, it considers whether the process, as something physical, has physical causal efficacy on physical circumstances beyond itself. This is why the Introduction was careful to talk about the causal efficacy of *conscious processes*, rather than about the causal efficacy of *consciousness*, i.e., of *being-conscious*. It comes to the same thing in the case of identity theories, but it could be that, under a non-identity theory, the conscious process as a physical entity has all the causal efficacy one could wish for whereas being-conscious is held to be denied any efficacy of its own (see section 6.1

on this "causal exclusion" problem, which has major differences from this paper's PLCD challenge).

One reason to highlight the identity/non-identity distinction is that the empirical and evolutionary considerations alluded to in the Introduction do not actually support the idea that consciousness itself has causal efficacy over and above the idea that conscious processes do. For instance, it is irrelevant from the point of view of evolution whether (i) consciousness itself is said to have causal efficacy or instead (ii) physical circumstances that are necessary and sufficient for consciousness are granted that efficacy. In a non-identity theory, if the causal efficacy of those circumstances helps to explain why those circumstances evolved in the organisms in question, it thereby helps to explain why consciousness evolved, because those circumstances cannot hold without consciousness being present, and consciousness cannot be present without those circumstances holding. In particular, the double-dissociation experiment mentioned in the Introduction does not supply more evidence for identity theories than for non-identity ones. Equally, under a non-identity theory we could still maintain the Facilitation Hypothesis, read as being about the causal efficacy of conscious processes rather than necessarily of consciousness itself.

3. The Physical-Level Causal Dispensability (PLCD) Challenge

Here I explain the core aspect of the challenge as advertised in the Introduction and then go on to the specific implications alluded to there. The core aspect is of philosophical interest in its own right, and challenges causal-pattern theories to take a stand on the causal-dispensability issue this paper raises. But it is in the specific implications that we see difficulties arising as regards how individual organisms need to be structured and as regards evolution.

3.1. The Core Aspect of the Challenge

The core aspect rests on the observation that (at least in existing causal-pattern theories) any causal effect of any process, conscious or not, on circumstances beyond[9] itself at some moment has, as a full cause, just some state B reached at that moment by the conscious process, plus possibly some enabling conditions arising elsewhere in the organism. In giving an account of such an effect Z, we do not need to consider the process's governing causation giving rise to B (the causation within the process and coming into it, giving rise to B). Any other pattern of governing causation that could have given rise to state B at the moment in question would have done just as well as regards the causal effects Z that B has.

Now, that argument was about causal effects arising from the process (on matters beyond itself) at any given moment, but we should also consider the causal effects of the process over its whole time course, arising from different states B at different moments. This matter is just an extension of the point already made—any pattern of governing causation that gives rise to the sequence of states B, across time, that the process's causation gives rise to would have led to the same causal effects. It is in this sense that the requirement that the theory at hand places on the nature of this causation, in order for the process to be conscious, is irrelevant as regards the causal effects of the process beyond itself.

Now, one may justifiably claim that the whole process also acts (together with any enabling conditions) as a full

[9] By "beyond" a process I mean outside it or subsequent to it. In using the term "outside" I do not imply there is necessarily a geographically separate part of an organism in which the process lies. A process might only involve certain aspects of physical state within the spatiotemporal region occupied by the process—e.g., only electromagnetic state, for the sake of example. So, a physical circumstance could be outside the process but within its spatiotemporal region.

cause of effects beyond itself. This way of talking is on a par with saying that a storm caused the destruction of a house even though just one component of the storm, such as a tornado in a particular place, destroyed the house. But the storm is a non-minimal full cause because it involves other aspects that can be removed—"dispensed with"—to get a smaller full cause. In particular, any causation of the tornado by other aspects of the storm can be dispensed with in providing a full cause. Equally, the causation within a conscious process and coming into it from outside itself, i.e. its governing causation, is dispensable.

Also notice that in saying that the storm as a whole caused the destruction of the house, there is no claim that aspects of the storm other than the tornado are themselves to be thought of as joint causes of the destruction along with the tornado. The whole storm is merely a convenient unit to mention, with the intent that some aspect of the storm was a full cause of the destruction. In particular, there is no claim that the causation of the tornado by other aspects of the storm is itself to be regarded as one of a set of causes that jointly cause the house's destruction. Equally, in saying that a conscious process caused something Z does not amount to a claim that its governing causation is one of a set of causes of Z. The governing causation is not an entity that in its own right is being granted causal power over Z.

So the point of our discussion is not that conscious processes as wholes, including whatever causation they contain, do not (or cannot be said to) causally affect circumstances outside themselves. The point is rather that there is a full cause of the effects that does not include that internal causation.

To fix ideas, we can take as a convenient and salient example a conscious process occurring inside a neural network. For simplicity, let's take the process to consist of the changing activation levels of the neurons in the network, plus the causation that produces the particular activation levels at each moment, with this causation arising from the signals that travel along connection fibres such as axons

between neurons. Here we ignore incoming causation for simplicity. Also, again for simplicity, let us assume that all causal effects of the network on matters beyond it are carried by fibres that come out of the network from some subset of the neurons in the network. Clearly, the only items that act as causes of the effects are, moment by moment, the activation levels of some neurons in the network. Nowhere in accounting for these effects do we need to refer to the causation in the network. In short, all the causal efficacy of the network on physical circumstances beyond itself is ultimately down to what is caused by the neurons' activation levels.

Now, if a theory takes being-conscious to be identical to some physical condition on a process that includes a requirement on its causal pattern, then the being-conscious is itself thereby immediately dispensable as regards the causal efficacy of the process. If the theory takes the non-identity route, it is still the case that being-conscious is instantiated if and only if the physical condition holds. Given that the causal pattern is dispensable in the above sense, then there is no warrant for saying say that the being-conscious can *not* be dispensed with. It would not make sense to claim that the being-conscious must be involved in an account of the causal efficacy of the process.

Finally, one might think that one could get round the dispensability of the causal patterning if one took the signalling along the inter-unit connections within the network as an additional part of state. It might be that the travelling of a signal along a connection could be part of a cause of something beyond the network. while also supporting the causation operating between units. But this just pushes the issue of causal patterning down to a more detailed structural level. First, if one took a connection to be a discrete series of nodes of some sort, with signals travelling between *them*, we just have a bigger network of roughly the same sort as before, leaving the issues above untouched. But I would submit that even if a connection is regarded as something truly continuous, the states lying continuously

along a connection at any given time are still just the results of causation (ongoing physical causation that "pushes" a signal along the connection), so that any effect arising from those states still does not have to be put down to the causation as such. If one did, on the other hand, say that the travelling of the signal (consisting, e,g, of changing electric field values along a fibre) was *itself* the *causation* between the original units, and not just a symptom of or carrier of the causing, while also being a matter of the changing state of the network, then one would actually be proposing that causation could be part of physical state and could contribute to causing just as other aspects of physical state can. But this is in essence what the meta-causal approach below does, so one is in fact meeting the challenge in broadly the way I suggest.

We should also notice that, of course, there can be *chains* of causation within the network and extending into and out of it. As a simple instance, state A can cause state B within the network, and B can in turn cause state Z outside the network. But this is very different from saying that, for instance*, the causing of B by A* causes Z or anything else. What directly causes Z is B by itself*,* not the causing of B by A, even though B is caused by A and even if B would not have occurred had A not occurred. What we *can* say is that Z is indirectly caused by A, by virtue of A's causing B. But this is not a matter of the-causing-of-B-by-A being the cause (or a cause) of Z. The upshot is that such chaining of causation does not alter the fact that it is only the states of the components of the network that cause anything, inside or outside the network.

3.2. Casual Dispensability in Some Illustrative Theories

Here I illustrate the core aspect of the PLCD challenge, and bring home its breadth of application to current consciousness theory, by looking briefly at some prominent but diverse approaches to consciousness. Analogous comments are, I believe, possible for many further theories.

We will see in section 3.3 that the specific-implications aspect of the challenge also applies to the three approaches.

One of the most prominent detailed theories of consciousness is the Integrated Information Theory (IIT: Oizumi *et al* 2014, Albantakis *et al* 2022; also Seth & Bayne 2022). As mentioned before, it is the main example of "causal-structure" theories in Doerig *et al (*2019), and indeed IIT discusses consciousness as entirely a matter of causal patterns, in a very complex way. IIT is therefore a good example for us, especially as typical examples of the nature of IIT are networks of units as in section 3.1. Thus the problem of causal dispensability for such networks applies directly to IIT.[10]

But two cautions are needed here. First, IIT is a probabilistic theory, and concerns the possible patterns of causation that a system can sustain, not just actual ongoing patterns, which are the focus of the present paper. Nevertheless, actual ongoing patterns are included within what is possible. Hence, it is reasonable to say that, according to IIT, a system is conscious in a certain period only if it provides a sufficiently integrated causal pattern over that period,[11] and the more integrated it is the more intensely conscious it is. The notion of integration has a precise, complex mathematical description, but the details are not important for our purposes.

The second caution is that, according to Cea and colleagues (Cea, Negro & Camilo 2023; Cea, Negro & Signorelli 2024) the question of what is real, and in particular what is physically real, is a vexed one in IIT, to the extent that those authors call IIT a "realist idealist" theory. There is even a question of whether the neural units in a neural

[10] The causal dispensability for ITT was also noted briefly by Cea (2023).

[11] The omission of an "if" alongside the "only if" here is deliberate, as we are omitting certain considerations.

network are to be regarded as having true physical existence. Nevertheless, it is fair to say that in most discussions of how IIT relates to, e.g., the human brain, most researchers take neural units and so forth to be straightforwardly physical.

The PLCD challenge also arises for another prominent theoretical approach, namely the Global Workspace approach (Baars 1988, 2017) whether in a neurally specific form (Dehaene & Naccache 2001, Dehaene & Changeux 2011) or not. See also Seth & Bayne (2022).[12] The approach's central feature on which consciousness depends is the global broadcasting of information amongst brain mechanisms via a central workspace. Although the theme of causation is typically not explicitly emphasized, the required broadcasting is implicitly a matter of suitably patterned causation, at least in part. One support for this stance is Seth & Bayne's (2022) statement that "GWTs account for changes in global states of consciousness in terms of alterations to the functional integrity of the workspace." Functional integrity here is a matter of causal integrity—a matter of the workspace working causally to serve broadcasting in the right way.

Suppose then that there is a conscious process in the system, so the right sort of broadcasting is happening. If the process causes something to happen beyond itself, that effect actually has as a full cause the states that have arisen from the broadcasting. These are states of the workspace itself, such as representations formed by integrating information from mechanisms putting information into the workspace, or states arising, because of the broadcasting, in mechanisms drawing information from the workspace. A full cause constituted along these lines does not amount to the broadcasting itself, along with its special causal pattern. Therefore, also, the process's being conscious is not part of that full cause. The points here are highly analogous to the ones made about IIT, only we are now talking at a level of description higher than the level of detailed networks of units, and of course the required type of causal patterning is different.

[12] Following Seth & Bayne (2022), I take the approach to be an approach to phenomenal consciousness as well as mere "access consciousness."

Our third illustration from prominent consciousness theorizing is provided by the higher-order thought approach (HOT theories: see Carruthers & Gennaro 2023 and Seth & Bayne 2022 for recent review).[13] My comments are confined to the actualist type identified by Carruthers & Gennaro (2023), as opposed to the dispositionalist type. The essence of the approach is that an organism is consciously entertaining a mental state if and only if: it has a higher-order thought about that state, and some further conditions, including about causation, obtain. An example of the former state, in the case of a person Sally, might be that *Sally hopes that she will be rich* [this would be called a first-order mental state], and an example of the higher-order thought might be that *she also believes that she hopes she will be rich* **[**a second-order mental state]. Sally would be entertaining the hope merely unconsciously if she had no second-order thought about it.

Crucially for us, as Carruthers & Gennaro (2023) make clear, there are constraints on how such a higher-order state *arises*. In particular: for the relevant lower-order state to be consciously held, the higher-order one must arise in a *suitably direct, internal, causal* and *non-inferential*[14] way from the lower-order one and other circumstances within the organism, and such causation must presumably be ongoing (Carruthers & Gennaro, *ibid.*), not merely to have obtained earlier. For example, for Sally *consciously* to hope that she will be rich, it is not enough for her to have the above second-order belief just because she infers it from her own behaviour (such as entering a lottery) or from other mental states such as a memory of Peter telling her that she hopes to be rich. On the other hand, it cannot be a matter of just a simple causal

[13] I believe it also affects the similar higher-order perception (HOP) approach (Carruthers & Gennaro 2023). It may even affect it more, to the extent that perceptions are yet more restrictedly tied by causation to what is perceived than thoughts are to what they are about.

[14] Whether only conscious inference is excluded or unconscious inference is also excluded is a matter of variation between theory versions, it would appear from Carruthers & Gennaro (2023).

link within Sally, because, according to the approach, the higher-order state does not always appear when the lower-order one does. So, altogether, consciousness is present only if a suitably constrained pattern of causation amongst the subject's mental states is present.

But the approach appears not to address the question of how the needed pattern of causation of the higher-order state actually enters into the causation of effects that one might think could flow from the presence of the consciousness. It appears that any effects causally arising from Sally's consciously hoping she will be rich would have, as a full cause, merely the two mental states mentioned, plus possibly some enabling conditions such as the person being sufficiently undistracted.

3.3. The Specific-Implications Aspect of the Challenge

The causal dispensability noted in section 3.1 as the core aspect of the PLCD challenge immediately creates *some* difficulty for explaining how consciousness-compliant causal patterns, and hence consciousness itself could have arisen in evolution. The causal pattern could not, at least as a feature taken by itself, have been specifically selected for. However, this does not close the evolutionary question, because, for instance, beneficial effects caused by conscious processes despite that dispensability could have been selected for, so the consciousness-compliant patterns could have arisen just as one of several alternative qualities that could have evolved.

It is therefore useful to consider the evolutionary question further. Let's continue with the scenario used above, where some state B that is reached by a conscious process is a full cause of some circumstance Z beyond the process. (We ignore any additional enabling conditions as above for simplicity.) The conscious process might be the activity in some neural network, or it might take a different form, but we will often refer to the neural case for definiteness. Obviously, if there were an alternative causal way for state B to arise in N, where that alternative way is not "consciousness-compliant"— not satisfying the causal-

pattern condition imposed by the theory—then Z would still be caused. Intuitively, the neural network or other mechanism in which Z can arise (henceforth "the Z mechanism") does not "care" how B was arrived at.

Let's further suppose that the Z mechanism is a lower-level mechanism on a route towards the organism's action effectors such as limbs, and that this mechanism never contains conscious processing. Suppose also that the Z mechanism's evolved function is to respond to action intentions generated by conscious processes in N. I will mainly stick to this action-intention scenario for definiteness, but an analogous discussion could proceed on the basis of other sorts of cognitive product.

Then, there is a worry is that the Z mechanism could be "misled" or "misdirected" into going into state Z even though N has not in fact been holding any conscious process. This would be disadvantageous to the organism if it is important for the organism's survival in its environment to be governed, to the extent possible, by its conscious action intentions (i.e., consciously produced ones) as opposed to unconscious action intentions (i.e., non-consciously produced ones). This might be because, for instance, conscious processing in the type of organism at hand involves more careful reasoning about relevant factors than unconscious reasoning in that sort of organism does.

In the following I will argue that there is a challenge to the causal-pattern theory at hand to explain why potential misdirection of that sort is not an obstacle to explaining the evolution of conscious organisms (unless of course the theory deliberately allows for or commits to a non-evolutionary story about consciousness). I say "a challenge" as opposed to a black-and-white claim that the theory cannot provide such an explanation. Below I will cover several ways in which the theory might try to minimize the actual importance of the potential misdirection. Since I am only trying to give the theory the benefit of the doubt, I do not need to explain those ways in detail or show that they are actually viable. The burden of proof would be on the theory to argue in favour of them if necessary.

The main question an advocate of the theory might ask is: would the possibility of misdirection actually arise, or arise to any important extent? Now, we can certainly imagine fanciful, rather unnatural ways in which a Z mechanism as above could be misdirected. For instance, if N, the mechanism potentially holding conscious processes, is a neural network and B an activation pattern over some of its neurons, an external "surgeon" agent could for some reason of their own zap the relevant neurons in such a way as to put them into state B. But this observation is surely not of much evolutionary significance—or at least hasn't been so far!

Aside from fanciful possibilities on such lines, a more serious first possibility to consider is that the sort of action intentions (or other cognitive products of interest) produced by conscious processing could not, even in principle, be generated by non-conscious processing. I know of no evidence that this is the case,[15] so will put the possibility aside. A more important possibility to consider is that, even though in principle the action intentions (etc.) could be generated non-consciously, nevertheless in a given species of conscious organism they are, as a matter of contingent fact, never (or only rarely) generated in that way. So, the misdirection rarely if ever arises in the organisms, and there is no practical PLCD challenge concerning the organism. It simply doesn't matter that states such as Z above are fully caused by aspects of conscious processes that don't themselves include the consciousness-compliant causation. The Z mechanism needs no special provisions to notice conscious as opposed to unconscious intentions: it's always or almost always causally prodded only by conscious ones.

One way such organisms might have evolved is as

[15] And much reason, of at least a methodological sort, to suspect that it is not, arising from decades of work in AI, cognitive psychology developing detailed computational models of how complex reasoning, decision-making, action planning, etc can be done, where no-one would seriously claim that an organism instantiating the model would have to be conscious, and where in some cases working robots have been built that do complex reasoning, etc. but give us no reason to think that they are conscious.

follows, even if consciously intended actions provide no inherent advantage in the external environment over non-consciously intended ones. The idea is that the organisms' ability to be conscious arose because it was adaptively important for reasons other than the generation or handling of action intentions (or whatever sort of cognitive product one is considering). There are many in-principle possibilities one might think of here. For instance, for all we know, conscious processing in general is less resource-intensive than unconscious processing that produces the same or similar cognitive products. For instance, if the theory is of the Global Workspace sort, an advocate might be able to argue that trying to do the sort of information integration achievable via a global workspace in a way that does not involve such a workspace would be more resource-intensive. If conscious processing is indeed beneficial resource-wise, then it may then have been an evolutionary side-effect that most action intentions in the organism ended up being consciously generated. To continue the Global Workspace example, perhaps the workspace and the broadcasting and integration it enabled arose in evolution before the complex production and handling of action intentions, and that processing of intentions was slotted into the already available workspace mechanism. Moreover, an account on these lines is compatible with the proposition that other species of organism might have evolved differently, and thereby be non-conscious species, while still having comparable cognitive processing for building and handling cognitive products such as action intentions. This could happen even for a species in the same world environment, given that the different species could have additional advantages and disadvantages, so that it is the overall package of features in a species that is adaptive as opposed to just individual ones separately. Even more so could it be that species in a distinctly different environment could have developed as non-conscious.

There is a qualitatively different, conceivable reason for conscious mechanisms to have evolved, without any implication that misdirection importantly arises. This is that such mechanisms, because of their intrinsic nature, are either

easier for evolution to create in the first place or easier for evolution to keep in place once created. That is, they are, respectively, intrinsically more evolvable or intrinsically more evolutionarily stable than non-conscious mechanisms with somewhat similar behaviour-generating capabilities would be. (This might in principle be, for instance, because mutations, DNA string recombinations or other evolutionary mechanisms are more likely to create or preserve conscious mechanisms than non-conscious ones, for all we know.) The conscious mechanisms may not provide more advantageous behaviour for any individual organism for this to be true. Again, to continue the Global Workspace example for the sake of argument, perhaps the neural circuitry that is needed to support a global workspace is intrinsically more evolvable or stable than non-conscious alternatives would be, even if a non-conscious alternative would have worked just as well and just as economically in any individual organism.[16] [17]

I also reiterate here that I am trying to give the benefit of the doubt to a theorist who wishes to minimize the importance of misdirection. I submit that the burden of proof is on such a theorist to develop the minimization suggestions in detail and show that they are viable.

The above minimization routes all assume that there could be (say) an action-handling, Z mechanism as above that is always or mostly prodded by conscious intentions

[16] I hope it is clear that I am merely using the Global Workspace examples for purposes of illustration. There is no suggestion that other sorts of causal-pattern theory couldn't come up with similar stories. I have no motive to support the Global Workspace approach.

[17] In fact, in other work I am developing my own meta-causal theory in a direction where the handling of pain and other discomfort is a main driver of the evolution of consciousness, but the evolutionary advantage is at least in part, and perhaps wholly, down to the meta-causal, conscious mechanisms being more evolutionarily stable than comparable non-conscious mechanisms would be, rather than being a matter of distinctive advantage to individual organisms. I have presented initial thoughts on this in talks (Barnden 2023, 2024).

arising as aspect B above of the conscious-processing mechanism N. But, in fact, there is a significant worry that, in an *individual* organism, at least in a relatively advanced conscious species, there is a real and important possibility of B states sometimes being generated by unconscious processing and sometimes by conscious processing. Thus, the Z mechanism could be causally affected by both, and the question then is how that mechanism is to distinguish them, should it be advantageous to the organism for it to be able to do so.

The possibility arises because there is good reason to think that complex cognitive processing that is not conscious occurs importantly in human beings (and, hence, it becomes a real possibility also in higher animals). Some evidence is provided, for example, by Dijksterhuis & Strick (2016), Hassin (2013), Ivy (2023), and Lau & Rosenthal (2011). It is also traditionally pointed out that while driving a car we may suddenly realise that we have been competently doing complex manoeuvres with no memory of having been conscious while doing them, so that it's a real possibility that we did them non-consciously.[18] Equally, it is a common observation that creative insights, seemingly involving complex cognition such as analogy formation and problem-solving, often happen precisely when we turn conscious attention *away* from the matter at hand, or even when we are "sleeping on it" (Dijksterhuis & Strick 2016). It may even be that certain types of complex cognition that we tend naively to think of as normally conscious are systematically done *better* in an unconscious way than in a conscious way. But all we need for the current discussion is that there is an important class of cases where cognitive products arise unconsciously as well as an important class where those products arise consciously.

This is not to say that a given product arises through the

[18] On the other hand, it would be wrong just to take it as certain that we did do the manoeuvres unconsciously. We may, for instance, have done them consciously in the normal way and forgotten. So, again, my discussion is in the spirit of raising challenges as opposed to fatal arguments.

same detailed cognitive steps (such as reasoning steps) when done consciously versus unconsciously. For instance, analogical processing in a creative act might proceed differently when conscious from when it is unconscious. Or, as mooted above, perhaps conscious reasoning is more careful than unconscious reasoning. Nevertheless, there is no reason to suppose that at least some products of cognitive processing (an analogy, a conclusion of reasoning, an action intention, …) are often the same whether they were produced consciously or unconsciously: you cannot tell just from looking at the product whether it was produced consciously or unconsciously.

Then, it is reasonable to consider the theory that, instead of the brain having, for instance one network in which conscious intentions arise and another in which unconscious ones do, both can arise in the same network N, but as a result of different processing. This could be advantageous from a resource point of view, for instance. The two lines can be arbitrarily similar or different in principle, subject to the former involving consciousness-compliant causation and the other not doing so. Clearly, in this setup we have just the conditions in which a Z mechanism as above could be misdirected by the unconscious action intentions, playing the role of state B above.

So, evolution would have to come up with special, additional measures to ensure that the Z mechanism can distinguish conscious cognitive products from unconscious versions of them. For instance, might N send a special signal to the Z mechanism when the processing has been conscious? This essentially requires that the N network be capable of at least a primitive form of self-reflection, of knowing when it itself has been conscious. I am not claiming that this is necessarily something difficult to evolve, but it certainly raises a challenge to any theory that proposes it to explain how it evolved. One would be proposing that even lower animals that performed actions with conscious intention can (unconsciously!) detect when they are conscious.

One might also wonder whether the Z mechanism could, in effect, analyse the causal patterns in N by analysing, say, the sequence of activation-level states in N, and thereby

determine whether they are consciousness-compliant. But it is far from clear how this would be done even in the relatively simple case of a global-workspace theory—where we are asking the Z mechanism to detect when suitable broadcasting is taking place—let alone in the case of a theory such as IIT with a very complex notion of the causal integration needed for consciousness, or even in an actualist HOT theory, where the Z mechanism would need to be able to detect what sort of causal path had produced the higher-order state.

But, it may be that Z could respond to simple side-effects of conscious versus unconscious processing. For example, perhaps conscious processing happens to produce more intense activation levels in the B state, or elsewhere within N, than unconscious processing does. There may be no particular reason in principle why it should do so, but it just happens to have evolved to do so, and the Z mechanism could still have evolved to pick up on this as a sufficiently reliable indicator of consciousness.

But, again, it is up to a causal-pattern theorist to develop and substantiate such possibilities.

In summary, the PLCP challenge does not prove any theory to be wrong, but challenges the theorist to explain, first, how the causal dispensability of the required causal patterns is not in fact an operational difficulty in a given individual organism of the type envisaged. For instance, is it that the theory claims that consciously-produced and unconsciously-produced cognitive products of the same sort do not arise in the same mechanism? Or that they do, but other mechanisms can tell whether those products were consciously produced or not? Or, conscious and unconscious products are produced for some good operational reason, but other mechanisms do not have to be able to distinguish them, perhaps because they do not inherently have any advantages over each other for the organism? Secondly, the theorist is challenged to provide a specific adequate story about the evolution of the claimed type of conscious organism (if biological) even though unconscious analogues may well have been possible, even in the same evolutionary niche. This second part of the

challenge is a specific addition to the already large, contentious problem of the evolution of consciousness.

The specific-implications aspect of the PLCD challenge as laid out in this subsection apply to the three illustrative approaches to consciousness discussed in section 3.2. Those approaches are largely informed by consideration of human-level consciousness, so that they are confronted by the issues in the present subsection surrounding relatively advanced cognitive processing and the possible availability of both conscious and non-conscious ways of producing similar cognitive products within one and the same mechanism in an organism.[19]

4. Meta-Causation

There is at least one way one might think of for avoiding the dispensability of causation discussed in section 3. This route is the meta-causal approach advocated by this paper, whereby the causation in a conscious process could *itself* be a physical entity that can serve as a (partial) cause, and in particular a cause of circumstances beyond the process. But mainstream approaches to causation that are relevant to consciousness simply do not allow for this possibility. Thus, it is not just that current causal-pattern theories have committed a straightforward oversight in not incorporating meta-causation. It is rather that they are strongly inhibited from doing so by prevailing conceptions of causation (and often this is through not explicitly appealing to any particular theory of causation). I will discuss these conceptions in section 4.3. First, in section 4.1, I give further intuitive clarification of meta-causation, and in section 4.2 I will elaborate on how meta-causation can help us dispense with causal dispensability and thereby meet the PLCD challenge.

---

[19] I am grateful to the anonymous reviewers for pressing me on just why the existence of non-conscious alternatives to conscious processing presents a problem, and for pressing me to mention the resource-usage point and the point about relative evolvability of ways of achieving a function.

4.1 Meta-Causation, Intuitively

The term "meta-causation" has (unfortunately) had a variety of disparate meanings within the causation literature, but I use it to mean what some have called "higher-order causation" (Koons 1998) or "iterated causation" (Ehring 2009, Kovacs 2021). Briefly, meta-causation comprises instances of causation where a cause and/or an affected item is itself an instance of causation. To put it another way, meta-causation is where causings *themselves* are entities with causal efficacy or causal influenceability. Of course, an instance of meta-causation, a *meta-causing* in other words, is a special case of a causing.

Meta-causation in my sense will probably sound unfamiliar to many if not all readers, and is only sparsely discussed in the mainstream literature on causation.[20] Nevertheless, it is a fairly intuitive notion. To take up the example mentioned in the Introduction, consider a sentence like *"John made Mary cry, and this angered her mother"*, where the speaker intends *"this"* to refer to the *"making cry"*, i.e., the causing-to-cry. It is this causing that is being said to cause the mother's anger. Thus the speaker is claiming an instance of meta-causation.

Now, the sentence could be meant or interpreted in other ways, such as by respectively intending or taking *"this"* to refer to Mary's crying, not the causing-her-to-cry. Under this meaning, there is straightforward chained causation (X causes Y and Y causes Z): John caused Mary's crying, which

[20] This sparseness is my own impression, but it is also noted by Kovacs (2021). Kovacs briefly mentions the idea of God continuously causing the physical causation in the world. This is one area of where meta-causation in my sense has seen concerted discussion, but it is not relevant to the present article. In contrast to meta-causation, the notion of meta-grounding has seen considerable discussion, in the literature on "grounding." See again Kovacs (2021), for example. The sparsity of attention to meta-causation is therefore perhaps rather strange, given that grounding is often thought of as a synchronic parallel to causation considered as a diachronic matter.

caused the mother's anger. I hope it is obvious that this meaning is importantly different from the meta-causal one. However, the fact of alternative interpretations is irrelevant for my purposes—my intent is to illustrate a possible thing one can say or think, not adjudicate how English sentences should be meant or taken in given contexts.[21]

One can generate common-sense meta-causal examples at will, such as those above and those in Ehring (2009). In fact, they are important in the legal and moral domains. For instance, if John causes Bill's death, this causing is what may get John into trouble with the law, and it might then be said that that causing meta-caused John to be arrested, etc. Of course, the fact that people may in effect talk or think about meta-causings no more implies their actual existence in the world than the fact that people may talk of Santa Claus implies his existence. The common-sense examples I have been giving are merely in service of clarifying the very notion of meta-causation.

In the John/Mary/mother example, we have "left-handed" meta-causation. This is because if we apply the template "X meta-causes Y" to the example, it is the X that is itself a causing (a causation instance). There is also right-handed meta-causation, where it is the Y that is a causing. An example is in the possible sentence *"Sally forced Paulina to make Bill go away,"* meant or interpreted to say that Sally forcibly caused Paulina's causing of John to go away. There is also an "ambidextrous" type where both X and Y are themselves causings. I take ambidextrous meta-causation to be both left-handed and right-handed, rather than neither-side-handed.

In discussions of ordinary causation (causation that is not meta-causation), a cause or effect can be complex and involve more than one component cause or component effect. For

[21] The ambiguity in the example highlights the importance of distinguishing meta-causation from chained causation, which does receive very extensive discussion in the causation literature. In this regard I find the label "iterated causation" in Kovacs (2021) and Ehring (2009) for meta-causation a distinctly sub-optimal one, as it smacks (to me) of chained causation, not meta-causation.

instance, smoking and having a certain bodily constitution might be claimed to form a complex cause with a complex effect of having both cancer and smelly clothes. The same is naturally the case with meta-causation, in that the cause or effect can contain more than one causing as a component. Also, the cause in a meta-causing might consist not just of causings but also ordinary circumstances in the world, and similarly for the effect.

A further complication is that the right-hand side of a meta-causing might involve not the creation of a causing as in the above examples but instead the modification of one that would already have been in place, or of the destruction of a causing. Intuitive examples would respectively be that Sally forced Paulina to *violently* make John go away, when she would otherwise have gently done so, and that Billy *prevented* Paulina from making John go away.

4.2 Meta-Causation Meets the PLCD Challenge

If there *is* indeed meta-causation in the world, the argument claiming a PLCD challenge is dissolved. The causings within the causal patterns that lie within a conscious process can now, as entities in their own right, potentially have direct causal effects on matters beyond the process.

For now, it may help to contemplate an analogy between the neural-net scenario in section 3.1 with the John/Mary/mother example above, where a state B of a network causes an effect Z beyond the network. In this analogy, the mother maps onto the Z mechanism, and the mother's being angry maps into state Z. Mary's crying maps onto state B. John's causing Mary to cry maps onto the inter-neural causation within the network that led to state B. In the intuitive scenario, we are taking John's causing of that crying to be a physical entity that is the cause of the effect on the mother. By the analogy, we are taking the inter-neural causation of state B to be a physical entity that is the cause of the effect on the Z mechanism. A slight variant on both sides of the analogy is that the cause side of the meta-

causation could be a joint cause, with one component cause being the John-causing-Mary-to-cry or the inter-neural causing of B.

Of course, in order to meet the PLCD challenge, the causings that act as causes (or effects) in meta-causation need to at some detailed physical level, unlike the higher-level everyday-life sorts of cause and effect we entertain in intuitive examples such as the John/Mary/mother one.

There may be many different detailed theories possible about how meta-causation could be involved with conscious processes, but my intent in this paper is merely to mention that a meta-causal framework may have the means to meet the PLCD challenge, whereas non-meta-causal causal-pattern approaches need to demonstrate some other way of meeting it. Note that for simplicity I assume that, if there is meta-causation, it is part of the physical domain, not something that is itself non-physical but can link aspects of the physical domain. To have meta-causation be non-physical in that way may be a viable proposal in itself, but I set it aside here.

A little extra care is needed in saying that the argument for the PLCD challenge is dissolved. Our considerations have only shown that the causation within an organism's conscious process can directly, meta-causally affect, in *some* way, matters outside the conscious process, such as in an action-intention-handling Z mechanism as above. We have said nothing implying that the ultimate specific action is one that is in fact consciously desired or intended by the organism (i.e., by its conscious process). The organism may consciously intend to swing its tail to the right, and the conscious process may start a causal chain leading to *some* external action being made, but we cannot conclude from the above that this action is to swing the tail to the right. We can only say that meta-causation *allows* it to be the case that a conscious process that contains an intent for the organism to do action X directly meta-causes circumstances that are outside the conscious process and that lead to the doing of X. Ensuring a tighter link between intent and action requires additional theory.

There is a further qualification. It is not enough for individual causings within the causal pattern in the process to have miscellaneous, independent causal effects on matters outside. For us to be able to say that the consciousness-compliant causation within the process has the efficacy, it needs to be that the overall patterned set of causings combine in concert with each other to form a complex cause that is the cause side of the meta-causing of some outside effect. There is an interesting possibility as to how this might be achieved, though it requires further investigation. It arises from the idea that causation is what provides structure diachronically in the physical world (complementary to synchronic structure, existing in the world at a given moment). Then, taking "pattern" to be synonymous with "structure," it is plausible to suggest that, the vague, generic notion of a consciousness-compliant "pattern" of causation can be precisified to mean *the set of meta-causings that tie together the causings into the putative pattern*. Importantly, these meta-causings are themselves elements in the pattern, so that *they* themselves are patterned just by meta-causation. Thus, unlike the case with ordinary patterns, the pattern is not the way in which some items that are conceptually different from the pattern are arranged in it—rather, the items that are patterned are themselves part of the pattern*ing*. We have patterning that patterns itself, to put it vividly. It is now the case that having a set of individual causings, more specifically meta-causings, as causes that affect matters outside the process does realize the idea that the pattern as such is having an effect, because the pattern itself *is* the individual meta-causings themselves, taken together.

4.3. Meta-Causation via Current Approaches to Causation?

This paper's argument that there is a PLCD challenge does not advert to any one of the many different approaches to what causation is (Ehring 2009, Gallow 2022, Kutach 2014; see therein for references to specific versions of the approaches). Equally, a causal-pattern theorist who is also an adherent of any of the approaches to causation could

potentially try to encompass meta-causation and thereby hope to avoid the problem.

However, at the start of the section I mentioned that current mainstream causation theories that are relevant for consciousness do not allow for meta-causation. In using the term "allow for" I am avoiding a claim that they definitely do not allow it. Rather, most mainstream causation theory does not even consider meta-causation, and the few approaches that readily already encompass or could encompass meta-causation have disadvantages as regards being applied in a consciousness theory.

Different current approaches to causation present different levels and types of promise for, or resistance to, encompassing meta-causation (let alone using ordinary causation and meta-causation in an adequate theory of consciousness). The following brief comments just hint at the issues, which merit extensive discussion elsewhere. My comments assume that we want causation to be an objective part of reality, not just something imposed subjectively by people, as in "projective" views of causation. The desire arises from seeing causation as a constitutive aspect of consciousness and taking consciousness to be an objective aspect of reality.

There is one type of approach to causation that can already naturally encompasses meta-causation, as indicated by Ehring (2009). This is where causal relata (causes and effects) are taken to be facts. Going back to our intuitive examples, the fact of Mary crying at some time would be caused by some fact about John. But this causing C can itself be regarded as a fact, so can itself be a cause. In our example, C would be the cause of the fact that Mary's mother is angry. This causing is then, of course, a left-hand meta-causing, and is itself a fact that could act as a cause. Right-hand and ambidextrous meta-causings are also straightforwardly possible. However, facts themselves would need to be of an entirely objective nature, and, at least in an identity theory, we would need to be happy with the idea of physical reality being (in part) *composed* of facts. I find both requirements difficult.

Another, related, type of approach is to take causation to be a relation between situations in a situation- theoretic

metaphysics of the world (Barwise & Perry 1983), and to allow situations to have causings as constituents. Then meta-causings are again a naturally arising possibility. Having reality be composed of situations may be more acceptable an idea than that it be composed of facts, and objectivity may be more plausible than with a fact-based theory.

A strongly related and more common approach is to take causal relata to be events. So, in our intuitive example, the event of Mary's crying is caused by some event involving John. If we then postulate that causings are events, John's causing of Mary to cry can itself be the cause of the event of Mary's mother being (or getting) angry. As may be clear from these comments, I am here taking events to include what are often regarded as processes, such as someone starting and then continuing to cry or being angry, not just punctate happenings at particular points of time. This view naturally encompasses causings to the extent that these are also taken to be time-extended rather than temporally punctate.

The view of causation as the passing of some physical marker or conserved quantity, such as momentum or energy, between world constituents of some sort presents more of an immediate problem. For meta-causation, we would need such passing to be itself something that could send out or receive the postulated type of marker or conserved quantity. For example, if what is passed in causings is momentum, then for a causing to be on the cause side of a meta-causing, that very passing would need to possess momentum, distinct from the momentum being passed, and some of this distinct momentum gets passed to the effect side of the meta-causing. However, perhaps some other view of what is passed in causing may be more amenable to meta-causation.

The last type of approach I discuss here is the type that explicates causation, partly at least, in terms of counterfactuals. (I include here approaches based on difference-making or the effects of possible interventions on a system.) Suppose we say John's shouting at Mary causes Mary to cry, and mean by it that John shouted at Mary, she then cried, and if John had not shouted at her (and suitable extra conditions X required by the particular counterfactual theory hold) then Mary would not have cried. Then for this

causing to be the cause in the meta-causing of Mary's mother being angry, with meta-causings also cashed out counterfactually, we would have to say something like: had it counterfactually not been the case that we have the counterfactual circumstance that if John had not shouted at Mary (and if X) then Mary would not have cried, (and if X^t) then Mary's mother would not have been angry. This is a counterfactual *about* the holding of another counterfactual. If this makes sense at all, it requires major additional complications in the theory of counterfactuals.

5. The Prior Motivation(s) for Considering Meta Causation

In this section I will explain my motivation for considering meta-causation, developed in work prior to discerning the PLCD challenge. I will call this "the prior motivation" for convenience. In fact it is a motivation of claiming that conscious process themselves are based internally on meta-causation, rather than merely saying that meta-causation helps with the causal effect of conscious process on other circumstances.

But before going into this it is useful briefly to mention another couple of reasons for at least considering meta-causation in theorizing about consciousness, and in particular for considering it to be important within conscious processes.

First, there is a consciousness-related philosophical area in which meta-causation is at least an implicitly important consideration, although I am not clear how widely the point is recognized. The area centres on the question of whether or not we can directly experience causings (see, e.g. Beebee 2009, Groff 2013, Mumford & Anjum 2011). Prime examples here are whether, in seeing a bird bend a branch through standing on it, we see the bird's causing of the bending; and whether, when performing an action, we experience the causation involved. Or we could consider whether someone can consciously perceive the very causation involved in John's causing of Mary to cry (as opposed to just perceiving symptoms of that causing such as

his shouting at her). If one could have such perceptions or experiences of one's own action, then a meta-causal possibility arises. Suppose that one takes the plausible view that seeing-X *itself* involves causation by X of happenings Y in the brain or mind. Then if X is itself a causing, the causation of Y is actually a left-handed meta-causing of Y by X. Whether or not one believes that we *can* in fact see or otherwise experience causings, such considerations show the potential importance of at least considering the issue of meta-causation in consciousness theory.

Secondly, suppose that section 4.2's meta-causal response to the PLCD challenge is taken to be plausible. This involves "external" meta-causation, i.e. meta-causation acting between the causation within a conscious process and circumstances beyond it (but still within the organism, possibly). But then it is only natural to propose that it can also be important *within* the process, that is, to propose "internal" meta-causation. Many readers may already have wondered about this. I have concentrated in this paper on the external case partly for simplicity. But similar considerations apply to causal efficacy of the process at some moment on subsequent physical activity in the process itself. And there are good reasons for considering such efficacy. For instance, it is reasonable to think that (the physical realizations of) our conscious intentions and thoughts can have a causal effect on subsequent states in the process, or on the way our conscious perceptual attention moves between different aspects of our environment. Moreover, it may be important that conscious thoughts preferentially have such internal efficacy, in contrast to unconscious thoughts arising also as part of the processing. So, if we are already proposing external meta-causation to subserve one sort of causal efficacy of the process, it is natural also to propose internal meta-causation to subserve another sort.

But there are limitations to the claim. One is that some of the reasoning in the previous paragraph is about relatively advanced, thoughtful sorts of consciousness, and may not be relevant to the more basic core of consciousness that, for instance, I have so far focused on in developing my own theory (Barnden 2020, 2022). Another is that the previous

paragraph leaves it open that meta-causation only comes into play in an intermittent way, when needed for specific purposes.

Now to the "prior motivation". It is a motivation for going much further than the claim just made about internal meta-causation, and thinking that *meta-causation is a crucial constituent of (the physical base of) all consciousness, at all times, in all types of conscious organism.*

This motivation does not involve the thoughts about causal efficacy in the present article. It arose from considering a basic form of consciousness that is often said to be "pre-reflective" (see, e.g., Miguens, Preyer & Bravo Morando 2016) and is claimed by adherents to be at the base of all consciousness. So, any conscious episode is founded on pre-reflective consciousness, but may also include more sophisticated, reflective aspects of consciousness. If an organism is at some moment phenomenally conscious merely pre-/non-reflectively, then its consciousness does not involve conceptualization, deliberation or thought in any ordinary sense. Because of this and because my use of "reflective" may different somewhat from that of other authors, I now find it convenient to use the term "non-conceptual" rather than "pre-reflective," though recognizing that "non-conceptual" puts aside deliberation, thought, etc., not just conceptualization taken narrowly. Non-conceptual processing excludes, in particular, "reflection" in the sense in which we can reflect on some issue, be it about ourselves or something else. The prior motivation has focussed on non-conceptual consciousness because of a desire not to include, as necessary features of consciousness, ones that might preclude lower forms of life from being conscious. However, I make no particular proposal or prediction about which non-human forms of life *are* conscious.[22]

[22] My use of "reflecTive" does not in itself include "refleXive" notions such as consciousness being aware of itself or of a separate self that has the consciousness, although of course it allows, as a special case, conscious reflection on one's consciousness or on oneself. On the other hand, other authors often appear to take reflective to imply reflexive (because they are thinking of the metaphor of reflection of light in a mirror), while taking pre-

The prior motivation itself starts with the idea that consciousness possessed by a conscious process cannot just be a matter of the sequence of ordinary, momentary physical states that the process goes through, ignoring the process's governing causation (its internal causation plus incoming causation). By ordinary physical states I mean states such as the positions of particles relative to some basis such as the location of the organism, the values of electromagnetic fields, etc., the rates at which such values are changing, and so on. It is highly implausible that two processes that happened to go through the same ordinary physical states but had markedly different governing causation would necessarily both be conscious if one is; or that, if both are conscious, they would necessarily have the same phenomenal characters as each other (conscious feelings, conscious perceptions, conscious thoughts, etc.). The easiest way of seeing this is as follows.

Suppose that the process occurs in a neural or computational mechanism that proceeds in discrete time, so that there is a state at time *t1*, a state at time *t2*, and so on. If the nature of the governing causation did not matter, one could get the mechanism to repeat that sequence of states over a later time interval just by forcibly reloading into it, at each time step, the state that occurred at the corresponding time in the original process. This reloading is just another form that governing causation could take. Furthermore, it would presumably not matter if, in the course of this activity of successive reloading, the units such as neurons or computer registers into which state is being reloaded were

---

reflective to keep the reflexiveness but not the conceptuality, etc. But in fact my approach ends up fairly close to their thinking, in effect. I argue in Barnden (2022) that, under the MDyn theory, all consciousness intrinsically includes reflexive non-conceptual consciousness as a base. This argument is radically different from existing arguments in the literature to the same conclusion (often couched in terms of "inner awareness"), such as the memory argument (Giustina 2022). However, the matter is not central to the current article.

replaced by new replicas of them. If the causation from state to successive state within the original process does not matter, it is difficult to see how the identity of the particular neurons, etc. matters—they might just as well be ones suddenly brought into play. Also, we can just eliminate all fibres, wires etc., that join the units together (assuming those connections do not themselves count as units over which process state is defined). These connections only serve to carry the causation, so to speak, but we are saying we do not care about the causation. Once we have arrived at this point, it is very difficult to maintain that we still have a conscious process. There is a just a series of states in units that have nothing to do with each other, other than happening to occupy the same positions—and it is not clear why even this should matter. But we *should* still have a conscious process if the nature of the governing causation is irrelevant. Hence, we have arrived at a contradiction[23].

I give further argument on these lines in Barnden (2014, 2020), and include the case of a continuous-time process. Altogether, while the considerations do not amount to a watertight proof that the consciousness of a conscious process cannot just be a matter of its state trajectory, ignoring the nature of its governing causation, they are at least suggestive. Hopefully, the reader will already believe, at least tacitly, that a brain, say, is conscious because of the way it *works*—the way that the momentary neural states *cause* further ones. For the remainder of this exposition, I am happy to leave it as a working assumption, at the very least, that the specific nature of the governing causation of a process matters to whether it is a conscious process or a non-

[23] The thought experiment in this paragraph is very similar to the one independently devised by Gidon, Aru and Larkum (2022). These authors similarly mention that one possible reaction to the experiment is to propose that (in our terminology) suitably-constrained governing causation must be present for consciousness to exist. However, their purpose for the experiment is different from ours, and they do not go on, as we do in this section, to propose that conscious processes must involve meta-causation.

conscious one.

The comments up to this point can be couched as saying that consciousness resides in genuine processes and not in pseudo-processes, where a genuine process is a process that includes its governing causation, and a pseudo-process (Dowe 2009) is intuitively just a state trajectory. A classic example of a pseudo-process is the shadow of a flagpole, with the shadow moving as the sun moves. The idea is that successive states of the shadow are not causally related. The causation of the successive states of the shadow from outside the shadow itself—by the sunlight being blocked by the flagpole—is analogous to the causation involved in reloading states into units in the above scenario. So, the above scenario could be summarized as saying that the reloading merely gives rise to a pseudo-process.

I submit also that it is plausible that, during the course of a conscious process, the governing causation matters temporally *throughout* the process. It cannot just be that all is required is that, say, a certain pattern of causation is present in an early stage of the process but later it doesn't matter what the pattern is, or that it only matters at isolated points during the process's time interval.

But we then get a big question: just *why* does the governing causation matter? Of course, it matters in the basic sense that the causation is what leads to the particular sequence of states the process goes through. But we've argued that this cannot be *all* that the causation is providing. One answer to the question would be simply to say that it is a brute metaphysical fact that consciousness is present if and only if a certain type of governing causal pattern is present. I cannot prove that this stance is incorrect. But, for one thing, it gives no clue as to why any particular causal pattern should have any specific significance over and above its providing a specific state trajectory. The following thoughts seek to eliminate the bruteness by seeing if we can say more about the nature of the needed causation, and also to keep the presence of consciousness is an entirely objective matter—it cannot reside in how some cognitive agent that is considering or otherwise reacting to the process consciously or unconsciously construes that process.

I suggest that a neat—but perhaps not the only possible—answer to the mattering question is that the governing causation *matters in its own right TO the process itself*, in a sense of mattering that does not involve any act of construal or other cognitive act by the process. I emphasize the "TO" to convey the idea that the process is in some sense *taking account* of its own governing causation, not just being driven along by it as any process is. It is taking account of it in broadly the same way it takes into account circumstances in the organism's environment that are important to the organism.

In Barnden (2014, 2020, 2022) I argue against the idea that this taking-account consists of representation, in any normally considered notion of representation (see, e.g., Shea 2018). The representational approach here would be each state that the process goes through contains a representation of some at least of the prior causation within the process; and the causation of each state includes causing the presence of the representation within the state. The representation might in particular include a representation of this very causation. One reason I repudiate representation here is that it is still a matter of major debate whether there is any account of representation that makes it a completely objective, construal-free matter, and, in particular, makes it free of the way we as theorists view things. I concur with McClelland (2020: 460 n.5)'s statement, "Debates around the naturalization of intentionality ... have no immediate end in sight," where I take intentionality to include representation, and naturalization to require objectivity.[24] Another reason

[24] McClelland goes on to be optimistic enough about the naturalization to aver that it is a route to follow in trying to naturalize consciousness. Be that as it may, it should hardly inhibit a search for other approaches. An alternative to representation that is often proposed is the relation of "acquaintance". Acquaintance has in particular been proposed as a basis for consciousness (see Giustina, forthcoming, some essays in Knowles & Raleigh 2019, and critical discussion in Gennaro 2016). Acquaintance is typically talked of as a conscious matter, so to explicate consciousness without circularity we would need a not-necessarily-conscious version of acquaintance. So, for the purposes of providing what we want we might propose that at each moment a conscious

is that I wish the taking-account to be simple enough for plausibly existing in simpler organisms than human beings, and I doubt that a representation of the governing causation of a process is anything but a complex matter.

As far as I know, no consciousness researcher has suggested that conscious processes contain the above sort of representation of causation. If they have not, then this tends to support my claim that such representation is not the right way to go. But it is useful in any case to have explicitly rejected the representational proposal. Once the idea has been raised that a conscious process should take account of its own causation, representation of that causation is a natural thing for theorists to propose if they think of representation as the way organisms take account of circumstances of importance to them.

Having put aside the representational proposal, what proposal can we come up with? Surely meta-causation is a natural contender. Once we have noticed that the notion of meta-causation exists, it becomes an obvious tool to reach for in allowing the governing causation of a process to matter in an explicit way to it, especially given that a normal way for physical things to matter to each other is for them to have causal interactions. The basic meta-causal proposal, then, is that some at least of the governing-causation up to a given moment during a conscious process meta-causally affects the state at that moment and possibly beyond. Such meta-causation is then a constitutive element of conscious processes. We thereby get a *meta*-causal-pattern theory, irrespective of any other elements that may be also required for consciousness.

It is not excluded that there are types of process other than conscious processes that also have some form of meta-causal

process is acquainted in some such sense with some of its prior governing causation. However, one of the standard objections to the notion of acquaintance is that is a mystery and itself remains to be naturalized. In fact, a topic I wish to address elsewhere is the possibility that meta-causation provides a way of naturalizing an organism's acquaintance with its own conscious states.

taking-account of—the explicit mattering of—own governing causation. The arrangement of meta- causation that I propose in my own MDyn theory (Barnden 2020) is a very specific arrangement of meta-causation, and other arrangements would not deliver consciousness.

More importantly, I am inclined to make the following observation about ordinary processes in everyday life, such as a train moving along. The fact that the configuration of matter at hand counts for us as a train may involve us in thinking (at least unconsciously) of the governing causation to some extent. The assumed governing causation may matter to us in judging that a train is present. But it just matters *to us*. I would say that a train is a train merely because we humans construe the relevant lump of matter and its causation as constituting a moving train. Certainly, an account of causation based on, say, passing of markers or conserved quantities as mentioned in section 4.3 may provide an objective, or at least a more objective, way of describing causation within and into the train. However, even if so, this causation does not explicitly matter TO the train, only to us. So there is no need to postulate any sort or arrangement of meta-causation as *constitutive* of the train, let alone that there is meta-causal explicit mattering of the train's causation *to the train*.

6. Discussion

6.1. Relationship to the Causal-Exclusion Problem

The PLCD challenge is partially analogous to, but also importantly different from, the long-standing "causal exclusion" problem concerning mental states, notably conscious ones, in non-identity theories (Robb, Heil & Gibb 2023; also Eronen & Brooks 2024). I will call that problem a "challenge" to such a theory. It is irrelevant whether the theory is a causal-pattern one or not. The causal-exclusion challenge is in rough summary as follows, for the case of conscious processes.

Assume that physical circumstances are entirely caused by other physical circumstances ("causal closure" of the

physical). Then, once one has said that some physical circumstance P is caused by the physical circumstances that realize some conscious process according to the theory at hand, there is (allegedly) no room for saying that the consciousness, as such, in the conscious process has causal efficacy as regards P. Or, at least, the is no room for that if one rejects redundant causation of P by both the consciousness and by the physical realization of the process, where such redundancy would have to be systematic and rife in the world.

The partial analogy to the PLCD challenge is that, in both challenges, the consciousness in a conscious process (or: its property of being conscious) is being excluded from having causal efficacy on physical circumstances beyond the process, or that it least is not needed in order to account for that efficacy. I'll therefore say that consciousness is claimed to be "dispensable" by both problems, to continue the terminology of this paper. And one could make the parallel a little closer if one regarded the consciousness-compliant causal patterns in a causal-pattern theory as lying at a higher level of physical description than the items that the causation links, though this higher level is still regarded as, say, a neurophysiological level.[25] Then, granted that, in a non-identity theory, consciousness is regarded as being at a higher level than the physical (or is at a special higher physical level than some physical level taken as basic and for which causal closure holds), both problems claim that consciousness, as a higher-level phenomenon, is in a sense dispensable.

But the essential difference is revealed by the fact that the PLCD challenge applies just as much to identity theories of the causal-pattern sort as to non-identity ones, whereas the causal-exclusion problem is confined to non-identity theories. And, even under a non-identity causal-pattern theory, the challenges are separate ones. The physical nature of a conscious process might have all the causal efficacy one might wish for, while the causal-exclusion challenge might well still be considered to be unmet, because in fact the latter

[25] But I do not myself take this multi-level view.

challenge already assumes that the physical realization of consciousness itself has all the desired causal efficacy. Conversely, meeting the causal-exclusion challenge may well leave the PLCD challenge in place. For instance, allowing redundancy of causation between the physical and mental levels would not automatically meet the PLCD challenge.[26]

Corry (2013) analyses the causal exclusion argument in the form normally considered and points out that it rests on a particular notion of physical causal closure that emergentist non-reductive physicalists (a type of non-identity theorist in this paper's terms) need not subscribe to. Such a physicalist takes mental properties to be physical in a high-level sense but irreducible to (though supervenient upon) more fundamental physical properties. This allows mental properties as physical causes, thus not conflicting with closure. I find this approach to the causal-exclusion argument highly persuasive as the right move for a non-reductive physicalist to take, but again it does not negate the PLCD challenge. Under a non-identity causal-pattern theory, if the consciousness-congenial causation—which lies at an ordinary physical level such as the neurophysiological or lower==is missing then consciousness is simply not present, so it is beside the point that irreducible causal efficacy is assignable to conscious mental states. Thus, the operational and evolutionary difficulties in section 3.3 above still stand.

6.2. Relationship to the Unfolding Argument

Doerig *et al* (2019) levelled a well-known, if contentious, "unfolding argument." The argument is against all theories that they designate as "causal-structure theories", with IIT as the main target. The unfolding argument is in support of a

[26] Robinson (2023) also mentions that even under an identity theory there could be difficulties with causal efficacy of the mental. The present paper's considerations are different from these worries as well.

claim that causal-structure theories are either, on the one hand, just false, or, on the other hand, unscientific because unfalsifiable by scientific experiments. I leave extended discussion of the argument to future work, partly because it has not taken meta-causation into account. The problem Doerig *et al* raise relies on the idea that we can simulate the operation of a network that has loops (something required by IIT in order to obtain consciousness) by means of a network that does not have loops (a feedforward network). The new network has the same input/output behaviour as the original one, but cannot be conscious according to IIT. This substitution is thus a special case of the present paper's consideration of alternative causal patterns achieving the same outputs (causal effects on matters beyond the process) as consciousness-compliant ones. But the present paper concerns the metaphysical, physical and evolutionary issues here, not the epistemological issue of whether we can use current empirical, scientific means to find out whether a given process is conscious.

6.3. Downwards Causation and Related Notions

A large topic, which must mainly be left for treatment elsewhere, is the relationship of meta-causation to a suite of related notions, including downward causation (see, e.g., G. Ellis 2016, Eronen & Brooks 2024, Gallow 2022, Paoletti & Orilia 2017, Robb *et al* 2023), formal causation (après Aristotle, see Gallow 2022), hylomorphism (again building on Aristotle, and achieving some popularity, see, e.g., Shields 2020, Yates, forthcoming) and the notion of a complex, whole system applying constraints on the behaviour of the parts (see, e.g., Juarrero 2023). I will just make a few preliminary comments here.

The topic of downward causation includes the notion that a complex whole at one level in the physical domain can have, in its own right as a whole, a causal effect on its parts, thought of as sitting at a lower physical level. It is possible that meta-causation can realize, or at least help to realize, this scenario in a particular way. Consider a complex whole as

consisting of parts together with their causal relationships, and with those relationships being at the level of the whole and not at the lower level occupied by the parts. Then one can imagine a meta-causing that has within its cause side some or all of those causal relationships, and that has one or more circumstances about parts within its effect side. Such meta-causings would then be one special way of realizing downward causation from whole to parts (not necessarily precluding other ways). In addition, the causal relationships defining the whole could meta-causally affect lower-level physical matters outside that whole. This would also be downward causation, and would be at akin to the sort of meta-causal effects discussed in section 4.2.

However, meta-causation is a much more general notion than meta-causally realized downward causation. There is no reason for meta-causings whose cause side is at a particular, relatively high level to be confined to going downwards—their effect sides might include circumstances at the level of the whole rather than the parts. Indeed, we could have upwards meta-causation, for instance from parts of a whole to causings at the level of the whole. Thus, meta-causation and downwards causation at best overlap: they overlap on the special case of meta-causally-realized downwards causation, but there could be other forms of downwards causation and there could be other forms of meta-causation.

We also observe that downwards causation in the sense of causation from a level above the physical down to a lower, perhaps physical, level might be realized as type of meta-causation if the cause at the higher level is a causing and/or the effect at the lower level is a causing. Klinge (2019) presents a panpsychist proposal that appears to involve the latter sort of downwards meta-causation.

Similar comments apply to the hylomorphic notion that the form of something can be important for its causal efficacy on other things, as well as just defining how its parts are arranged. With regard to meta- causation, much the same considerations apply as with downwards causation. Let's suppose that the form of an entity, especially if it is processual as opposed to a static object, can reside in the causings between parts or aspects of the entity. Then the causation within or constituting the form might meta-causally

affect something (within the same entity or outside). Thus, meta-causation might supply or help to supply one version of causation-by-forms. However, there are other types of form to consider, not least geometric form, which is central to Yates (forthcoming); and equally not every case of meta-causation need be a matter of a form causally affecting something.

Similar comments again apply to the notion of the high-level nature of some system constraining lower-level matters, notably within that same system. If (some of) the causation involved in the system is viewed as a high-level matter, then there is room for proposing that that causation can meta-causally affect low- level matters, thereby constraining them. But much as with the causal efficacy of form within the previous paragraph, this does not exhaust the topic of constraints, and meta-causation need not be confined to high- to-low constraining.

On the other hand, to the extent that it would be legitimate to regard all causing as a form of constraining, a meta-causing would be a special case of constraining, with the interesting feature that its cause or effect side would itself involve constraining. In Barnden (2020) I suggest an instance of this strategy, in order to fit my own MDyn theory into an eternalist conception of the universe where there is no time-flow. The resulting eternal "block" then contains a web of constraining, some of which is the meta-constrai

ning of the block's constraining.

6.4. Fit with Modern Physics

An important objection to the idea of meta-causation is that it may appear to be a bad fit with modern physics. First, even putting meta-causation entirely aside, it has long been contentious to include causation at all in a physical picture of the world, especially at the basic physical level of particles and fields (Frisch 2023). But even if one is happy, as I am, to regard the equations of physics as describing something about ordinary causation, one might think there is no room to amend them to encompass meta-causation. Modern

physics is excellent and amazingly complete and accurate at describing and predicting what happens in the world. However, this point begs a question, because modern physics does not describe or predict consciousness, and to my knowledge the mentioned completeness and accuracy has not been experimentally demonstrated to apply within the relevant parts of a currently conscious brain. This point is the nub of responses that, for instance, dualists can make to arguments from physics. See for example Cucu & Pitts (2019). Part of the problem is that it is not known with any certainty which parts of the brain, and which physical aspects and levels (molecular, intra-neural, inter-neural connectivity, etc.) in those parts, are crucial for consciousness. And, as regards the meta-causal approach in the present paper, and even the specific MDyn theory in Barnden (2020), they are not at the stage of producing hypotheses as to what those parts or levels are. The theory leaves it open that the meta-causal effects could be at a detailed quantum level of description (for instance, see speculative suggestions in Barnden (2020) on meta-causation being involved in the dynamics of "objective collapse" in a quantum mechanical theory), or involve the fine detail of chemical transmission at synapses or electromagnetic fields between neurons, for example. So it is difficult to know whether any existing results from the wealth of detail neurophysiological explanation that has been done bear for or against the presence of meta-causation.

I would also submit that this paper's argument and the "prior motivation" for basing consciousness on meta=causation (see section 5) provide a justification for at least investigating the possibility of meta-causation, even if it looks foreign to current physics.

These comments are a rather programmatic response to the argument from physics, but I would also point to the mathematical formulation of aspects of the MDyn theory in Barnden (2020), involving "guard" conditions in laws. In that paper I discuss some ways in which a complex physical context could be accessed by such guards, and could, so to speak, release existing physical laws to act in unusual ways that explicitly refer to causation, and thus effect meta-causation. This could be the beginning of a path to seeing

how meta-causation and consciousness might reside in the context of (say) a brain while not doing so elsewhere in the world. Recall also that there may be forms of meta-causation that do not provide consciousness, and these may exist in appreciable quantities in unusual types of non-conscious system.

6.5. Future Work: Power Theory and Process Theory

I do not place strong restrictions on what a (conscious or non-conscious) process is. I view a physical process (the focus of interest in this paper) just as (a) some continuous trajectory of states over some non- zero time-interval, each state being the whole of, or some aspect of (e.g., the electromagnetic aspect of), the physical state of some spatial region, where the progress of the trajectory conforms to applicable laws of nature, plus (b) the causation that provides that progress. The overall spatiotemporal region occupied by the process can be one that is humanly defined rather than one that is objectively carved out in some way in the world. However, further development of the ideas in this paper may find it convenient to use a more restrictive notion of process.

The work in this paper has been influenced in a general way by work on process theory and work claiming the fundamental importance of dynamism and activity, such as work by Röck (2022), Seibt (1990, 2013) and Zubiri (2013). In particular, some deep similarities to, but also deep contrasts to, the activity-based work of Fichte (1982) dating from around the turn of the 19th century are briefly explored in Barnden (2020). As regards detailed theorizing about consciousness, this paper aims to contribute to the contemporary body of work that rests on processuality or activity rather than static conditions. For diverse examples of such work, see Gennaro (2012), Oizumi, Albantakis & Tononi (2014), Strawson (2017: especially chapters 1,3,8) and Van Gulick (2006). The paper is oriented towards causation being ontologically fundamental and a crucial constitutive component of processes.

There are many potential lines of future work, but some

are to tie this paper's ideas in more specifically with research in the (mutually overlapping) areas of process theory and power theory, given that this paper's causal realism meshes with some ideas in these areas. For instance, the relationship to the work of Ingthorsson (2021) is of special future interest, and causation is a major concern in power-theoretic metaphysical approaches (see, e.g., Groff & Greco 2013, Jacobs 2017, Mumford & Anjum 2011). However, causation is involved in various disparate ways in different approaches of this type (as is evident from comparing, say, Bird 2013, Buckareff & Hawkins (2023) and Mumford & Anjum) and it is difficult to make a general statement about the prospects for the approaches to encompass meta-causation. The view of causation in the MDyn theory (Barnden 2020) overlaps that of Mumford and Anjum (2011) to some extent. However, there is a tension with Mumford and Anjum's idea of causation being a matter of the cause turning into (i.e., becoming) the effect in an "unfolding process" but not being something that is over and above cause and effect [*ibid.*, p.119]. I may need precisely to reify that unfolding process as something over and above the original cause and original effect.

I should caution that, while B. Ellis (2013) discusses what he calls "meta-causal powers," these are not powers that somehow provide meta-causation in this paper's sense. Rather, his powers are called causal powers, and meta-causal powers are ones whose manifesting can destroy, create or modify other causal powers. So, in short, the exercising of his meta-causal powers is not a matter of exerting causation on causings, but only on powers to cause. In light of this, the term "meta-causal power" should be parsed as "meta-[causal power]" not "[meta-causal] power," and Groff's (2013) term "meta-power" is preferable from my point of view.

7. Conclusion

This article has addressed a broad class of theories of (occurrent phenomenal) consciousness, ones that give a crucial role to patterns of causation in the question of whether a process is conscious or not. The paper has argued

for two main claims. The first is that, since the causation in a conscious process can in principle be replaced by differently patterned causation without thereby affecting how the process causally affects other circumstances within the organism, the theory is challenged to show why this does not create an operational difficulty within individual organisms of the type of envisaged, and is challenged to show why a story about the evolution of consciousness can still be given. The second main claim is that the challenge can be met if meta-causation is brought in, so that the causation in a conscious process can in itself have causal effects beyond the process (and also within the process itself).

Existing causal-pattern theorists may understandably feel that effort should go into developing their own theories, as opposed to switching riskily to a radical idea such as meta-causation. My reaction to this is that there is value in both lines of research, and in confronting each line with any insights developed while following the other.

The paper went on to summarize the author's previously published motivation for considering meta-causation. This motivates taking meta-causation to be a crucial, ever-present part of all conscious processes, in whatever organism, not just a way for conscious processes to interact with the rest of the organism. The motivation arises from the idea that a conscious process should take some sort of explicit account of its own governing causation, i.e., the causation composing it combined with causation coming into it. In effect, the present paper presents extra support for bringing meta-causation into consciousness theory. Conversely, the prior motivation, in bringing in meta-causation for one purpose, naturally lends support to it being proposed for other purposes also, such as its purpose in the present paper.

Although this paper focusses on causation within the physical domain, some of the considerations might be generalizable to causation outside that domain. For instance, the idea that a causal pattern may lack causal efficacy of its own—it cannot itself be a cause—could arise in views of causation in non-physical realms, or causation that spans

physical and non-physical realms. This point could be relevant to a theorist who wishes to develop a detailed causal-pattern theory of consciousness in a non-physical realm, and would therefore be aided by including meta-causation. This point holds even if evolution is not a concern there, because aspects of the PLCD challenge are about the organization of processing in individual entities irrespective of considerations of evolution.

Finally, the paper fits with a highly "non-Humean" view of causation, that is, a view that regards causation as a real matter in its own right, and not just a way of couching regularities that exist across the universe without explaining how those regularities themselves arise. In fact, the motivation for proposing meta- causation in section 4 is deeply connected to an argument put forth by Hawthorne (2004) and refined by Toby Friend (2022). This argument uses considerations of consciousness to argue against a Humean view of causation. I hope to explore the detailed connections elsewhere. But the noteworthy commonality for present purposes is the use of the topic of consciousness—viewed as esoteric by many, and as postponable till we know more about physics, the brain etc.—as a basis for arriving at a fundamental proposal about the deepest nature of the physical world, not just a conclusion about consciousness. (It does this in a different way from that in which panpsychism (Brüntrup & Jaskolla 2016) places consciousness as the very foundation of the physical world.)

Author contributions: All authors contributed to the discussion and writing of the manuscript.

Funding: No funding agency.

Disclosure Statement: No potential conflict of interest was reported by the author.

References

Baars, B. (1988). *A cognitive theory of consciousness.* Cambridge: Cambridge University Press.

Albantakis, L.; Barbosa, L.; Findlay, G.; Grasso, M.; Haun, A.M.; Marshall, W.; Mayner, W.G.; Zaeemzadeh, A.; Boly, M.; Juel, B.E.; *et al.* (2022). Integrated Information Theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *arXiv* 2022, arXiv:2212.14787.

Baars, B.J. (2017). The global workspace theory of consciousness: Predictions and results. In S. Schneider & M. Velmans (Eds.), *The Blackwell Companion to Consciousness* (2nd Ed.), pp. 227242). Hobokon, NJ: Wiley-lackwell.

Barnden, J.A. (2014). Running into consciousness. *J. Consciousness Studies, 21* (5–6), pp.33–56.

Barnden, J.A. (2020). The meta-dynamic nature of consciousness. *Entropy 22*(12), 1433; https://doi.org/10.3390/e22121433 .

Barnden, J.A. (2022). Pre-reflective self-consciousness: a meta-causal approach. *Review of Philosophy and Psychology***,** online 22/1/22; DOI: 10.1007/s13164-021-00603-z .

Barnden, J.A. (2023). The meta-causal theory of phenomenal consciousness: Evolutionary adaptiveness despite simulative zombies. Talk at *First Web Conference of the Intl Society for the Philosophy of the Sciences of the Mind* (ISPSM 2023), 24/25 Nov & 1/2 Dec 2023. {Talk slides available on request.]

Barnden, J.A. (2024). Evolutionary implications of the meta-causal theory of phenomenal consciousness. Talk at *Joint Session of the Aristotelian Soc. and the Mind Assoc.*, University of Birmingham, UK, 12-14 July 2024. [Talk slides available on request.]

Barwise, J. & Perry, J. (1983). *Situations and attitudes.*

Cambridge, Mass.: MIT Press.

Beebee, H. (2009). Causation and observation. In H. Beebee, C. Hitchcock & P. Menzies (Eds), *The Oxford Handbook of Causation.* pp.471–497. Oxford: Oxford University Press.

Ben-Haim, M.S., Dal Monte, O., Fagan, N.A., Dunham, Y., Hassin, R.R., Chang, S.W.C. & Santos, L.R. (2021). Disentangling perceptual awareness from nonconscious processing in rhesus monkeys (*Macaca mulatta*). *PNAS, 118*(15), e2017543118, https://doi.org/10.1073/pnas.2017543118.

Birch, J. (2022). The search for invertebrate consciousness. *Noûs, 56*(1): pp.133-153.

Bird, A. (2013). Limitations of power. In R. Groff & J. Greco (Eds), *Powers and Capacities in Philosophy: The New Aristotelianism,* pp.25–47. New York and London: Routledge.

Brown, S.A.B. (2022). How much of a pain would a crustacean "common currency" really be?. *Animal Sentience 32*(23).

Brüntrup, G. & Jaskolla, L. (Eds) (2016). *Panpsychism: contemporary Perspectives*. Oxford Scholarship Online, October 2016.

Buckareff, A.A.& Hawkins, J. (2023). Emergent mental properties are not just double-preventers. *Synthese.*

Carruthers, P. & Gennaro, R. (2023). Higher-order theories of consciousness. In E.N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), https://plato.stanford.edu/archives/fall2023/entries/conshigher

Cea, Ignacio (2023). IIT's fundamentality of consciousness: its problems and an emergentist alternative. Talk at *First Web Conference of the Intl Society for the Philosophy of the Sciences of the Mind* (ISPSM 2023, 24/25 Nov \& 1/2 Dec 2023.

Cea, I., Negro, N. & Camilo, M.S. (2023). The fundamental

tension in Integrated Information Theory 4.0's realist idealism. *Entropy, 25,* 1453, https://doi.org/10.3390/e25101453 .

Cea, I., Negro, N. & Signorelli, C.M. (2024). Only consciousness truly exists? Two problems for IIT 4.0's ontology. *Frontiers in Psychology, 15:* 1485433, doi: 10.3389/fpsyg.2024.1485433.

Cleeremans, A. & Tallon-Baudry, C. (2022). Consciousness matters: phenomenal experience has functional value. *Neuroscience of Consciousness, 2022*(1), DOI: 10.1093/nc/niac007 .

Corry, J. (2013). Emerging from the causal drain. *Philosophical Studies, 165*. pp.29–47. DOI 10.1007/s11098-012-9918-3 .

Crump. A. & Birch, J. (2022). Animal consciousness: the interplay of neural and behavioural evidence. *J. Consciousness Studies, 29*(3-4), pp.104-128.

Cucu, A. & Pitts, B. (2019). How dualists should (not) respond to the objection from energy conservation. *Mind and Matter, 17*, pp.95–121.

Dehaene, S. & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*, 200227. https://doi.org/10.1016/j.neuron.2011.03.018 .

Dehaene, S. & Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition, 79*(1), pp. 1–37.

Demarest, H. (2017). Powerful properties, powerless laws. In J.D. Jacobs (Ed.), *Causal Powers,* Chapter 4. Oxford: Oxford Scholarship Online.

Dijksterhuis A. & Strick, M. (2016). A case for thinking without consciousness. *Perspectives on Psychological Science 11*(1): pp.117--132.

Doerig, A., Schurger, A., Hess, K. & Herzog, M.H. (2019) The unfolding argument: Why Integrated In-

formation Theory and other causal structure theories cannot explain consciousness. *Consciousness and Cognition, 72*, pp.49–59.

Dowe, P. (2009). Causal processes. In *Stanford Encyclopedia of Philosophy,* Spring 2009 edition.

Dung, L. (2023), Tests of animal consciousness are tests of machine consciousness. *Erkenntnis,* published online Nov 2023, https://doi.org/10.1007/s10670-023-00753-9 .

Ehring, D. (2009). Causal relata. In H. Beebee, C. Hitchcock & P. Menzies (Eds), *The Oxford Handbook of Causation.* pp.387–413. Oxford: Oxford University Press.

Ellis, B. (2013). The power of agency. In R. Groff & J. Greco (Eds), *Powers and Capacities in Philosophy: The New Aristotelianism,* pp.186–206. New York and London: Routledge.

Ellis, G. (2016). *How can physics underlie the mind: top-down causation in the human context.* Dordrecht: Kluwer.

Eronen, M.I. & Brooks, D.S. (2024). Levels of organization in biology. In E.N. Zalta & U. Nodelman (Eds),

The Stanford Encyclopedia of Philosophy (Summer 2024 Edition), https://plato.stanford.edu/archives/sum2024/entries/levels org-biology/.

Fichte, J.G. (1982). *The Science of Knowledge: With the First and Second Introductions.* Edited and translated by Peter Heath and John Lachs. Cambridge, UK: Cambridge University Press.

Friend, T. (2022). Why I'm not a Humean. *Pacific Philosophical Quarterly* (2021), pp.1–23, DOI: 10.1111/papq.12398 .

Frisch, M. (2023), Causation in physics. In E.N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), https://plato.stanford.edu/archives/win2023/entries/causation-physics/.

Gallow, J.D. (2022). The metaphysics of causation. In E.N.

Zalta & U. Nodelman (Eds), *The Stanford En- cyclopedia of Philosophy* (Fall 2022 Edition), https://plato.stanford.edu/archives/fall2022/entries/causation- metaphysics/.

Gennaro, R.J. (2012). The consciousness paradox: consciousness, concepts and higher-order thoughts. MIT Press.

Gidon, A., Aru, J. & Larkum, M.E. (2022). Does brain activity cause consciousness? A thought experiment. *PLoS Biology 20* (6): e3001651. https://doi.org/10.1371/journal.pbio.3001651.

Giustina, A. (2022). A defense of inner awareness: The memory argument revisited. *Rev. Phil. Psych., 13*(2), pp. 341-363.

Giustina, A. (forthcoming). Inner acquaintance theories of consciousness. In *Oxford Studies in Philosophy of Mind.*

Groff, R. (2013). Whose powers? Which agency? In R. Groff & J. Greco (Eds), *Powers and Capacities in Philosophy: The New Aristotelianism,* pp.207–227. New York and London: Routledge.

Groff, R. & Greco, J. (Eds) (2013). *Powers and capacities in philosophy: the New Aristotelianism.* New York and London: Routledge.

Hassin, R.R. (2013). Yes It Can: On the functional abilities of the human unconscious. *Perspectives on Psychological Science 8*(2), pp.195--207.

Hawthorne, J. (2004). Why Humeans are out of their minds. *Noûs, 38*(2), pp. 351–358.

Ingthorsson, R.D. (2021). *A powerful particulars view of causation.* New York and London: Routledge.

Ivy, S. (2023). Unconscious intelligence in the skilled control of expert action. *J. Consciousness Studies,* 30(3--4), pp.59--83.

Jablonka, E. & Ginsburg, S. (2022). Learning and the

evolution of conscious agents. *Biosemiotics, 15*, pp.401437.

Jacobs, J.D. (Ed.) (2017). *Causal powers.* Oxford: Oxford Scholarship Online.

Juarrero, A. (2023). *Context changes everything: how constraints create coherence.* MIT Press.

Kirk, R. (2005). *Zombies and consciousness.* Oxford: Clarendon Press (Oxford University Press).

Klinge, F. (2019). The role of mental powers in panpsychism. *Topoi 39*, 1103–1112 (2020), https://doi.org/10.1007/s11245- 019-09632-x.

Knowles, J. & Raleigh, T. (Eds) (2019). *Acquaintance: New Essays*. Oxford, UK: Oxford University Press.

Kolodny, O., Moyal, R. & Edelman, S. (2021). A possible evolutionary function of phenomenal conscious experience of pain. *Neuroscience of Consciousness, 7*(2): niab012.

Koons, R.C. (1998). Teleology as higher-order causation: a situation-theoretic account. *Minds and Ma- chines, 8* pp.559-585.

Kovacs, D.M. (2021). The question of iterated causation. *Philosophy and Phenomenological Research,* published onine in 2021. Also as: *Philosophy and Phenomenological Research, 104*(2), pp.454–473. DOI: 10.1111/phpr.12782

Kutach, D. (2014). *Causation.* Cambrige, UK: Polity Press.

Lamme, V.A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences, 10*(11), 494-501.

Lamme, V.A.F. (2022). Behavioural and neural evidence for conscious sensation in animals: An inescable avenue toward biopsychism? *J. Consciousness Studies, 29*(3–4), pp.78–103.

Lau, H. & Rosenthal, D. M. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences 15*(8), pp.365–373.

Levin, J. (2023). Functionalism. In E.N. Zalta & U. Nodelman (Co-Principal Eds), *Stanford Encyclopedia of Philosophy,* Summer 2023 Ed., https://plato.stanford.edu/archives/sum2023/entries/functionalism/. Department of Philosophy, Stanford University, Stanford, CA 94305.

Ludwig, D. (2022). The functional contributions of consciousness. *Consciousness and Cognition, 104,* 103383. DOI: 10.1016/j.concog.2022.103383.

Marchetti, G. (2022). The why of the phenomenal aspect of consciousness: Its main functions and the mechanisms underpinning it. *Frontiers in Psychology, 13*: 913309. DOI: 10.3389/fpsyg.2022.913309

Miguens, S., Preyer, G. & Bravo Morando, C. (Eds), *Pre-Reflective Consciousness: Sartre and Contempo- rary Philosophy of Mind.* Routledge.

McClelland, T. (2020). Self-representationalist theories of consciousness. In U. Kriegel (Ed.), *The Oxford Handbook of the Philosophy of Consciousness,* pp.458–481. Oxford, UK: Oxford University Press.

Mumford, S. & Anjum, R.L. (2011). *Getting causes from powers.* Oxford University Press.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83*(4), pp.435–450.

Newen, A. & Montenmayor, C. (2023). The ALARM theory of consciousness: a two-level theory of phe- nomenal consciousness. *J. Consciousness Studies, 30* (3-4), pp.84–105.

Niikawa, T., Miyahara, K., Hamada, H.T. & Nishida, S. (2022). Functions of consciousness: conceptual clarification. Neuroscience of Consciousness, 2022(1), DOI: https://doi.org/10.1093/nc/niac006

Oizumi, M., Albantakis, L. & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS*

Computational Biology 10(5): e1003588.

Paoletti, M.P. & Orilia, F. (Eds), (2017). *Philosophical and Scientific Perspectives on Downward Causation.* London: Routledge.

Robb, D., Heil, J. & Gibb, S. (2023). Mental Causation. In E.N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Spring 2023 Ed.), https://plato.stanford.edu/archives/spr2023/entries /mental- causation/.

Robinson, W. (2023). Epiphenomenalism. *Stanford Encyclopedia of Philosophy,* Summer 2023 Edition, https://plato.stanford.edu/archives/sum2023/entries /epiphenomenalism/ .

Röck, T. (2022). *Dynamic realism: Uncovering the reality of becoming through phenomenology and process philosophy.* Edinburgh: Edinburgh University Press.

Schaffer, J. (2016). The metaphysics of causation. . In *The Stanford Encyclopedia of Philosophy*, Fall 2016 edition, ed. Edward N. Zalta. http://plato.stanford.edu/archives/fall2016entries/causationmetaphysics/, accessed 1 December 2016.

Seibt, J. (1990). *Properties as processes: A synoptic study of Wilfrid Sellars' nominalism.* Atascadero, CA: Ridgeview Publishing Co.

Seibt, J. (2013). Process philosophy. In E.N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2013 ed.), http://plato.stanford.edu/archives/fall2013/entries /process-philosophy/.

Seth, A. & Bayne, T. (2022). Theories of consciousness. *Nature Reviews: Neuroscience, 23*, pp.439--452.

Shea, N. (2018). *Representation in cognitive science.* Oxford: Oxford University Press.

Shields, C. (2020). Aristotle's psychology. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), https://plato.stanford.edu/archives/win2020/entries

/aristotle-psychology/

Strawson, G. (2017). *The Subject of Experience.* Oxford Scholarship Online, March 2017.

Van Gulick, R. (2006) Mirror mirror — is that all? In K. Williford & U. Kriegel (Eds), *Self-Representational Approaches to Consciousness*, pp.11–39. Cambridge, UK: Cambridge University Press.

Yates, D. (forthcoming). Hylomorphism, or something near enough. In D. Yates & A. Bryant (Eds), *Re- thinking Emergence.* Oxford University Press.

Zubiri, X. (2003). *Dynamic structure of reality.* Translated from the Spanish and annotated by Nelson R. Orringer. Champaign, HE: University of Illinois Press (Hispanisms series).