Article

The Meta-Dynamic Nature of Consciousness

John Barnden [1]

[1] School of Computer Science, University of Birmingham, B15 2TT, UK; jabarnden@btinternet.com; (+44)(0)121-440-5677

Abstract

How, if at all, consciousness can be part of the physical universe remains a baffling problem. This article outlines a new, developing philosophical theory of how it could do so, and offers a preliminary mathematical formulation of a physical grounding for key aspects of the theory. Because the philosophical side has radical elements, so does the physical-theory side. The philosophical side is radical, first, in proposing that the productivity or dynamism in the universe that many believe to be responsible for its systematic regularities is actually itself a physical constituent of the universe, along with more familiar entities. It also proposes that instances of dynamism can themselves take part in physical interactions with other entities, this interaction then being "meta-dynamism" (a type of meta-causation). Secondly, the theory is radical, and unique, in arguing that consciousness is necessarily partly constituted of meta-dynamic auto-sensitivity, in other words it must react via meta-dynamism to its own dynamism, and also in conjecturing that some specific form of this sensitivity is sufficient for and indeed constitutive of consciousness. This leads to a proposal for how physical laws could be modified to accommodate meta-dynamism, via the radical step of including elements that explicitly refer to dynamism itself.

Keywords: consciousness; pre-reflective self-consciousness; physicalism; dynamism; causal productivity; meta-causation.

1 Introduction: A Theory and its Philosophical and Physical Sides

Elsewhere ([2–4]) I have been developing a distinctive approach to the problem of how phenomenal[1] consciousness can be explained as an aspect of the physical world. The theory has been largely philosophical in tone up to now, but the present article is a preliminary step toward giving it some precise mathematical, theoretical-physical scaffolding. I use the label MDyn for the theory for ease of reference in this article. The label stands for the *Meta-Dynamic* theory of consciousness.

The reader is warned that, because the theory makes a radical proposal on the philosophical side, it also requires radical proposals on the physical-theory side. In addition, the philosophical side, partly because of its own radical nature, is very much still in development, and will remain so for the forseeable future. The physical-theory side is therefore highly provisional and subject to change. And, as this article represents a first attempt to formalize the philosophically ideas in physical terms, the formalization is formally elementary and schematic in many places. It is a really a framework from within which to launch further development.

A radical element on the physical side of the theory involves proposing modified forms of existing laws of physics, as couched in, for example, Schrödinger's equation in quantum theory. However, the intention is that under "normal" circumstances—for instance, outside systems such as human brains in

[1] I will normally suppress this word—in this article consciousness will always be phenomenal consciousness, where there something that it is like to be in the state.

the course of conscious experience—these laws will reduce to working in their original way, or at least doing so closely enough for divergences to have been undetectable. There is a further highly radical step as regards what the extra parts of laws are *about*. I return to this point later in this Introduction.

As regards the radical step of proposing any sort of modified laws, let alone for the ambitious purpose of accommodating consciousness, I should mention that there have been plenty of examples of modification proposals within physics. One class of examples is provided by collapse models or collapse theories in quantum theory (see [21,33] and references therein). These seek to modify existing physical equations so that they account for events of "collapse" of the quantum wave function (into particular alternative physical possibilities that it can be interpreted as jointly representing) as just part of the overall dynamical unfolding of the universe. This is contrast to the more traditional route of treating them as a separate and rather mysterious type of event that is not itself explained by the dynamical equations of physics. The Bohmian approach to quantum theory (as represented for example in [7]) not only dissolves collapse into the general dynamics (where it is effectively replaced by bifurcations of trajectories), but also explicitly raises the issue of how quantum theory might be yet further modified, as will be briefly outlined below. I therefore see my own proposal on the physical side as of a piece with such work, to the extent that my modifications to laws are structurally similar to those of, say, [33] and [7]. Interestingly, both those works engage with aspects of the question of how to reconcile consciousness and physics. Because of the link to consciousness of the Bohmian approach as expressed in [7], I will use this work as my reference point for the approach, following here [43]. However, both approaches are also important from the point of view of explaining how (what is called) collapse works—or more properly, providing an objective, precise basis for the intuitive, approximative notion of collapse—without any necessary connection to consciousness.

This mention of collapse models is merely to paint the current paper as not being just out on a lonely limb, *NOT* because my approach itself relates consciousness to collapse. Readers who know of proposals that have been made about the connection of consciousness to quantum theory should note that this paper's proposal is neither (i) on the lines that conscious observation has a role in causing collapse events to happen (see [38] for a recent discussion) nor (ii) one that casts consciousness as being constituted by collapse events (as in the Orch OR theory, [26]). Rather, it relates consciousness instead to the non-collapse side of physics, though it is not actively averse to giving collapse a role should this become needed.

Physical laws apply everywhere and always, let us assume. The idea mentioned above that a new feature in a law might in practice only have a significant effect in very special physical circumstances is part of the "conditionality response" that Cucu and Pitts [10] discuss. This response is a rebuttal of objections to forms of dualism concerning consciousness that propose that the physical realm interacts with a non-physical realm where consciousness resides. However, their arguments, while focussed on objections to such dualism, extend more broadly, and support the legitimacy of proposing disturbances to normal physics inside conscious systems, whether the disturbances involve just a modified physics or (also) interactions with a non-physical realm. Cucu and Pitts cover especially the question of such disturbances implying breakages of physical symmetries, and hence local losses of conservation of energy, momentum, etc. I leave to further research the question of whether this breakage/conservation-loss issue affects MDyn negatively; but in any case Cucu and Pitts argue, plausibly in my view, that one should not regard that issue as an overriding concern.

In the remainder of this Introduction, and in the next section, I merely summarize the motivations and key tenets of the philosophical side, without trying to argue for them. In the Appendix, I include arguments and some other details for readers who may be interested.

The philosophical side of MDyn rests, in part, on a particularly radical version of the "anti-Humean" idea that the regular, apparently law-governed self-consistent unfolding of the universe is not there through pure arbitrariness. So, according to anti-Humeanism, it is not just a coincidence that regularities in, for example, how electrons and magnetic fields behave, are the same here and now as those a second ago and a foot away. Some anti-Humeans have claimed that, instead, there is some sort

of productiveness that, so to speak, pushes the universe forward in a systematic and consistent way. In short, things happen because they *made to* happen (or, to allow for fundamental stochasticism, because they are in a set of things, one of which must happen). Some have called this pushing-forward the [metaphysical or causal] "oomph" in the universe (see [14,35,48] for suggestions and discussion).

I adopt this oomph notion, though I use the term *dynamism* for it. At the same time, I consider that previous proposals about it have not taken it seriously enough physically as opposed to just metaphysically. By metaphysically here I mean as something that is somehow behind the universe but need not be taken account of by detailed physical theory. My radical anti-Human stance is, in part, precisely to say that, if we are going to propose oomph at all, why not *propose it as a genuine physical aspect of the universe, one that needs to be accounted for by mathematical physics*—no less genuinely physical than items mentioned in existing versions of physics, such as fields, spacetime curvatures, etc. To put it in terms of ontology or of what is real, oomph should be included in the list of things that a completed physics takes to be real.

I take this *reification of dynamism* yet further, and propose that one should take dynamism to be a first-class citizen of the universe, in the sense that it can interact with ordinary physical aspects of the world (electromagnetic fields, curvatures, etc.) rather than just be the pushing-forward that supports the interaction of the latter aspects with each other. I call any unmediated interaction of dynamism with something a *meta-dynamic* interaction. Such interacting is itself part of the dynamism of the universe, and can itself interact with physical items. As a special case, I claim that instances of dynamism can interact with other instances of dynamism, even themselves, not just with "ordinary" physical items such as those I have mentioned.

To come finally to consciousness: I propose that *consciousness is a matter of meta-dynamism.* In particular, a conscious process must be *meta-dynamically auto-sensitive*: directly sensitive to the very dynamism that holds it together and makes it a genuine process. That auto-sensitivity is the meta-dynamism. This characterization of consciousness in terms of meta-dynamism appears to be a novel proposal, though see [2] for a small number of links to possibly strongly related proposals. Although I have brought consciousness in after discussing meta-dynamism, it was in fact considerations about consciousness that led to proposing meta-dynamism and therefore needing to propose dynamism as a whole as a first-class physical citizen. These considerations about consciousness derive indirectly from a notion of pre-reflective self-consciousness in the phenomenal and neo-phenomenal tradition, as will be outlined below.

For intuitive and heuristic reasons it is useful to think of dynamism as a form of "causation" at the fundamental physical level, though this usage of the term "causation" departs greatly from usages in most philosophy of causation, as I explain in the Appendix. Then, meta-dynamism becomes a form of *meta-causation*, i.e. where instances of causation themselves are causes and/or effects in their own right. Meta-causation can be entertained as a possibility whatever type of causation one has in mind. For instance, the notion can be used commonsensically in sentences about everyday matters—consider *"Don's causing democracy to collapse caused Vlad to admire him"*. (Of course, such usages of the notion do not imply that, objectively, there is real meta-causation, least of all at the fundamental physical level.) Interestingly, meta-causation has seen remarkably little discussion overall, even in the philosophy of causation, as Kovacs [32] notes, let alone the particular form of meta-causation in this article, namely meta-dynamism. Thus MDyn is doubly unusual: not only does the theory make the novel move of linking consciousness to meta-causation, but also, meta-causation itself is an unusual topic aside from any link to consciousness.[2]

---

[2] Some occasions of the notion being discussed are cited in the Appendix, Section A.2. Unfortunately, muddying the waters, the term meta-causation has separately been used in a variety of ways that are distinctly *not* analogous to my meaning, including to mean "downwards" or "top-down" causation, usually from the mental to the physical [18]. Also, see A.2 for the point that in [2] I rejected the word "causation" and did not use "dynamism," preferring instead a different term.

On the physical-theory side, MDyn is radical not only in proposing modifications to law-expressing equations in the form of added terms, factors, etc., and allowing for entirely new laws, but also in that the proposed additions and new laws (if any) are centrally *about* spatiotemporally located instances of dynamism. This is the physical side's version of the philosophical notion that dynamism is a first-class physical citizen.

Furthermore, these dynamism instances are, inherently, *temporally (and spatially) non-local*, so that a law that is partly about such instances is not explicitly about merely one (generic) point in spacetime in the conventional way. Existing laws generally state how quantities constrain each other at a particular instant, even though a quantity may be a rate of change of something with respect to time and so, highly implicitly, may concern other times. Adlam (2018) discusses various ways in which temporal non-locality has been suggested in physics.

As a result of the views adopted, MDyn is, to an enormous degree, "multiply realizable" at least in principle. For instance, the claimed meta-dynamism does not rely on any particular type of matter, such as biological matter, and does not rely even on realization in any particular physical "flavour" of the universe, such as its electromagnetic side, its gravitational/curvature side, etc. Potentially, the right interactions might exist in arrangements of many different specific types of physical matter, field, etc. The theory doe not even rely on consciousness being within an object in the everyday sense, as opposed to, say, within diffuse, "disembodied" fields. I am not making any particular claim here about where consciousness may reside, but merely pointing out the liberality of the theory. It may *in practice* turn out, for reasons for reasons of stability and survival, that consciousness can only successfully reside in tangible objects, and that these objects will tend strongly to be composed of a particular form of matter in some range of temperature, pressure, etc. This article does not address these pragmatic matters.

The latitude does not go so far as necessarily to imply panpsychism [8], i.e. some form of the idea that consciousness exists throughout the universe, even possibly within individual elementary particles or other small ubiquitous physical systems. I will discuss opportunities for MDyn to confine consciousness to certain regions or structures. This possibility exists partly because MDyn (probably) requires some *special* form of meta-dynamic auto-sensitivity, not just any such. The special form might only arise under special physical circumstances. The other side of this coin is that there may well be important and perhaps widespread forms of meta-dynamism that have nothing at all to do with consciousness, or are mere precursors to consciousness in evolution, say. The present article does not make definite conjectures about such forms, but they would be a good target for future research.

A related consideration is that, while MDyn conceives of [meta-]dynamism as being at a *deep* physical level, that does not necessarily mean that it is (only) about a *microscopic* scale. Just as quantum theory is not confined to the microscopic scale in principle, some meta-dynamism might intrinsically exist at a large scale. Another related point is that a theory such as MDyn can aspire to be what one might call *bathypsychic* (a term I introduce in [2])—in other words, characterizing the conditions needed for consciousness at a deep physical level, rather than postulating conditions that intrinsically involve high-level entities—without also being panpsychic. While panpsychic theories tend strongly to be bathypsychic, for good reasons, we should not assume a bathypsychic theory must be panpsychic.

The multiple realizability extends to computational artefacts, as long as their implementation in a physical device provides suitable meta-dynamism. According to MDyn, it is not enough just to run a sufficiently complex and cunning algorithm in a conventional computational substrate such as a present-day computer. The dynamism in the transitions between the physical realizations of the computational states needs to have a direct, meta-dynamic effect on the computations (or on other such dynamism). A consequence of this is that if one ran a computer simulation of a conscious system, a conscious brain for example, and even if the simulation covered the minutiae of meta-dynamism, the result would not be conscious unless the simulation algorithm itself were physically implemented in a suitable meta-dynamic way. And, even if it were implemented in a way that amounted to consciousness, that consciousness might not be the one simulated: it would be at an entirely different

level of description of the overall system. In terms of thought experiments such as the Chinese Room ([12,49]), MDyn agrees that the mere bandying about of symbols in, e.g., the Chinese Room is not itself sufficient for consciousness. However, it does not use this as an argument against the possibility of a conscious symbol-processing system, as in principle the system might be realizable in the needed meta-dynamic way. This does mean that the system would not purely be one of symbol-processing: it would also process dynamism itself, so to speak. But in practice, conscious systems of any sort, symbol-processing or otherwise, will usually need *some* interaction with the physical world. MDyn just leads to a very specific, deep sort of interaction.

The structure of the remainder of the article is as follows. Section 2 summarizes the main assumptions, tenets and stances on the philosophical side of MDyn. It also states the overall nature of a central argument behind MDyn, leading to a claim that meta-dynamic auto-sensitivity is necessary for consciousness. Section 3 presents some considerations informing the transition from the philosophical side of MDyn to the physical-theory side.[3] Section 4 presents the current preliminary steps to give the theory a precise mathematical clothing. I stress that I aim only to formulate the very foundation of what one needs in terms of physical laws to cope with the theory, *not* a mathematical theory of consciousness at anything remotely like the human level. Section 5 engages in discussion and further remarks. It points out in particular that the theory naturally accommodates the possibility of a very pure form of conscious experience that is nothing more than an experience of being aware of its own continued existence (cf. notions of pure consciousness, and, relatedly core and minimal consciousness, that a variety of other consciousness researchers have entertained or discussed [13,20,39,40,45,54,56,58]. The section also briefly discusses possibilities for cross-fertilization between MDyn and three other theories, though without any necessary implication that the other theories would thereby adopt meta-dynamism as a necessary feature of consciousness. The theories covered are the Integrated Information Theory [42] of consciousness, based on complex patterns of causation (but not meta-causation), the collapse model of Kremnizer and Ranchin [33] based on a related but now fundamental-physical notion of Quantum Integrated Information, and the Orch OR theory [26] that casts consciousness as being constituted by quantum-wave collapse events. Section 6 briefly concludes.

2 Philosophical Assumptions, Stances and Tenets

I first lay out some general stances, assumptions and claims about consciousness and that do not specifically touch on distinctive features of MDyn. Next I mention a key, fundamental assumption about consciousness that MDyn makes. It concerns *pre-reflective auto-sensitivity*, which is a weakening of a much-discussed notion of pre-reflective self-consciousness (or pre-reflective self-awareness). Then I give the overall shape of an argument that goes from PRAIS to the main claim of MDyn, a Necessity Claim that says that consciousness must involve meta-dynamism appropriately. I then present a Sufficiency Conjecture and some adjuncts to it, to the effect that suitable meta-dynamism is sufficient for and constitutive of core consciousness.

The argument mentioned and various other are matters are expressed in some detail in the Appendix: section A.5 for the argument, sections A.1–A.4 for the other matters.

2.1 Consciousness and Genuine Processes

- I will leave it open whether consciousness is *purely on/off* (it is either present or absent, with no gradations, at a given place and time), *purely graded* (i.e., it comes in degrees that can be anywhere in some interval up from zero), or *impurely graded* (i.e., a combination of on/off and

[3] Sections 2 and 3, while somewhat lengthy, have been put in the briefest form consistent with given the reader adequate understanding of the main ideas and their motivations and contexts, given their unusual nature.

graded, in that the conscious degree is either zero or at least equal to some fixed positive value). For simplicity I leave out the possibility that gradations could be in steps rather than being continuous. My own suggestion is that consciousness is impurely graded, but nothing in the article depends on this suggestion. The possibility of on/offness will be a significant issue in the Discussion section.

- Consciousness is a property, in the first instance, of *processes*, rather than of static states such as beliefs, or of brains, people, etc. These other types of thing can be conscious in a derivative sense. A conscious process is an episode of experienc*ing*.
- The processes that have consciousness are *entirely physical.* Being-conscious is a *physical property*.[4]
- Implicit here is an emphasis on consciousness being an *objective* property. A given process is or is not conscious, or is conscious to some degree, as a matter of objective fact, ultimately definable in purely physical terms, and not requiring any observer to construe the process as being conscious.
- A conscious process does not need to come equipped with a separable something that can be called a self, and "I" or a subject of consciousness, although relatively rich episodes of consciousness might be usefully regarded as involving (or being related to) such a thing. Rather, the experiencing is itself the "self" of the experiencing. This is a so-called *non-egological* view of experiencing, at least when we considering basic forms of experience. It motivates my use of the prefix "auto-" when others would use "self-".
- The notion of a process in general (conscious or otherwise) must take account of its possibly being something that, intuitively, moves as a unit through space. For instance, a process in a car engine or brain moves around in space as the person moves around. So in general the spacetime region occupied by a whole process over time will have a very complicated, wiggly shape through spacetime, if one thinks of the latter as a block, even if at any given time it happens to be spatially simple.
- When a process can be viewed as something that is moving from place to place, this cannot affect whether it is conscious or not. And the absolute location of a process in spacetime cannot matter to whether it is conscious or not: all that matters is its internal activity and (possibly) the interactions with the physical world outside itself.
- However, I consider it in principle possible for a conscious process to have no interactions with anything at all outside itself, even if this is unlikely to occur in practice.
- The concept of a process in MDyn allows both for cases where a process is viewed as being on some substrate such as a brain, with the process viewed as different from the substrate, and cases where no division is considered: there is just activity, though it may be that some of this activity happens to be, say, movements of matter. In the latter case, the activity may include the activity of what would be regarded as the substrate under the former view.
- When a process is viewed as being on a substrate, then I assume that the particular identities of components of the substrate are unimportant: replacement of functionally identical components does not affect the nature of the process, and in particular does not affect whether it is conscious or not. (This assumption connects to thought experiments such as those concerning replacing the neurons of a conscious brain by other neurons or by artifical, identically operating components. NB: in MDyn, though, the functional equivalence has to extend to the meta-dynamism.)

---

[4] Having said this, MDyn could be generalized. If one thought there was a non-physical realm, and one also thought it had something analogous to physical dynamism, then one could propose transplanting MDyn into that realm, and using a notion of consciousness that encompassed processes in that realm as well as physical processes. This observation is not peculiar to MDyn. A theory such as IIT, even if intended as physical, might be similarly broadened.

2.2 The Assumption of Pre-Reflective Auto-Individuating Auto-Sensitivity

MDyn is crucially based on an assumption of *pre-reflective auto-individuating auto-sensitivity* (PRAIS). This is the assumption that *any conscious process is continuously sensitive to its own existence as a physical, conscious process, in a pre-reflective way, where the auto-sensitivity allows the process to identify which process it itself is.* That is, the process, at each moment, detects its own presence as separate from other activities, where this detection is pre-reflective (or pre-conceptual) in the sense of not involving intellectual matters such as concepts, beliefs, thoughts or reasoning (in normally accepted meanings of these words).[5]

Note carefully that this assumption expresses only necessity. It does not say that PRAIS is by itself sufficient for consciousness.

The term "continuously" in the assumption demands that the auto-sensitivity should exist at *every moment* in the time-span, because any break would lead to the process being more properly viewed as only intermittently conscious. (I am putting aside the possibility here that the break could be literally instantaneous, rather than covering an interval of time.) So the question of consciousness would be pushed down to break-free portions. Thus, the remainder of this article is implicitly about uninterrupted conscious processes. Note that a process with interrupted consciousness that is also capable of entertaining beliefs might falsely believe itself to have been continuously conscious. The auto-sensitivity discussed in this article does not require an intermittently conscious process to notice the absence of its own consciousness at previous times. At most, it requires the process, in effect, to notice (possibly unconsciously—see below) that its own consciousness has been present when in fact it has been.

Apart from building in an assumption that consciousness is physical, PRAIS is a weakening of an assumption that *pre-reflective self-consciousness* (or pre-reflective self-awareness) is an inherent part of all consciousness. This assumption goes back at least to phenomenologists such as Husserl and Sartre, and has been recently much discussed by neo-phenomenologists and commentators (see e.g., [19,24,34,50,55,59], wherein also historical links can be found). The central claim is that in [phenomenally conscious] experiencing, that experiencing is automatically also, in some way that may be difficult to articulate, experiencing that experiencing. The self-consciousness does not automatically amount to what we normally regard as conscious introspection, which proponents of pre-reflective self-consciousness would say is something that involves doing the introspection by concept/thought-imbued reflection. Indeed, pre-reflective self-consciousness is held to be the reason that introspection can be phenomenally conscious in the first place: the reflectiveness does not itself confer this property.

The key difference of PRAIS from pre-reflective self-consciousness is that the auto-sensitivity is not *assumed* to be conscious. (See the Appendix section A.2 for the reason for this weakening.) But in fact the arguments in MDyn do ultimately lead to the conjecture it will, at least intermittently, amount to pre-reflective self-consciousness. This point will not explicitly be argued in the present article.

The *auto-individuating* aspect of PRAIS is an addition new to this article, though it was implicit in previous developments. Indeed, it always been implicit in the notion of pre-reflective self-consciousness, and indeed implicit in the use of the very term "self." (See also the discussion of "ipsiety" in [59].) It has always been at least tacitly assumed that it is not enough for an episode of experiencing (a conscious process) just to be conscious of the experiencing in a way that is undifferentiated from its consciousness of the overall activity in the world that that experiencing is part of. Rather, the experiencing has got to be consciously sensing itself as an entity separate from the activity environment it is embedded in. To use a metaphor, the experiencing is a current in the activity of the world that is not just aware of the movement of the ocean it is in, and therefore only implicitly aware

---

[5] Because of the pre-reflectiveness, MDyn readily *allows* for consciousness in tiny infants and non-human creatures, going perhaps quite a way down in the biological monarchy. However, I do not make specific claims about where in the monarchy consciousness actually lies.

of itself in some weak sense, but aware of itself as one particular current within that ocean. Similarly, in PRAIS as opposed to pre-reflective self-consciousness, that auto-individuation must again exist, though now it is not assumed to be a conscious matter necessarily.

2.3 From PRAIS to the Necessity Claim

Before the argument proper to be made, it is necessary to argue (see Appendix section A.4) that one cannot define consciousness just in terms of of *trajectories of ordinary state* (e.g., electromagnetic or positional states) that a process goes through. Rather, one needs the states to be causally linked with each other, using some objectively real relation of causation. (At this point, I am not bringing in my particular notion of reified dynamism, but proceeding at a more general level.) The intuitive crux of the arguments on this point is that, without bringing in the causal binding, a suitable series of completely unconnected states could, highly implausibly, have to be accepted as conscious: e.g., a succession of states each of which is in a different person's brain.

Now, turning to PRAIS, one might think that the auto-sensitivity could readily be achieved just by having each state in the process hold some sort of representation of some or all of the prior states and their causal binding (though explicit representation of the latter might arguably be dispensed with). So, part of the causation of a state as the process unfolds includes the causation of the updating of that representation. The representation could be of one of the many forms proposed in theories of representation in cognitive science, AI, etc. (see, e.g., [51] for comprehensive survey and discussion of the fundamental space of possibilities). But I argue (see Appendix section A.3) that there is no theory of representation that is *completely* objective—free of *any* subjective construal of what is represented by what—and *completely* pre-reflective. So representation cannot be a basis of an objective attribution of PRAIS.

But if extant theories of representation cannot work for the purpose at hand, what can? The overall problem with representation is the "distance" from the representatee to the representation, allowing scope for non-objectivity and reflectiveness to intrude. So one might propose that the past nature of the process (relative to a given moment in it) *directly* affects the current state, with no physical mediator at all. That is, one would be relying on a bold move of allowing temporal non-locality into physics (see [1] for a range of possible forms of temporal non-locality), where for instance a present state can in part be caused directly by a state in the past, rather than through a continuous unfolding of state in the interval since that past state. But one option here will not work: the option of just having past ordinary states, either individually or as a collection, and not including their causal binding, have a direct causal affect on the current state. This cannot work because the current state could be "deceived" by an additional, bogus past history (trajectory of states) that was gerrymandered (faked) to be isomorphic to the process's actual history and also to affect the current state as the same way as that real history, but *not* to contain the causal binding that would naturally be needed for that history. As explained further in the Appendix, Section A.5, the result is that the current state cannot be held objectively to auto-individuate, so that we get a contradiction with the PRAIS postulate.

This is where MDyn suggests that the desired auto-individuating auto-sensitivity would be achieved if *causation* up to the current state—as opposed to prior ordinary states—were to have a directly causal effect on that state There would be causation *from that causation itself* to some aspect of the current state: i.e., a form of meta-causation. One could also put this as saying that the meta-causation would achieve a new sort of representing of past process history that *is* entirely objective and pre-reflective.

If there are no other better suggestions of how PRAIS might be achieved, this meta-causal suggestion stands as a promising and perhaps even the best explanation. Then, observing that ultimately the causation that is detected and the causation involved in the detecting must rest on causation at the basic physical level—partly because otherwise the meta-causation itself could be just

a matter of optional construal, rather than objective—we get the suggestion that fundamental-level meta-dynamism will achieve PRAIS.

Notice that the reification of dynamism was proposed in MDyn *in order to* fulfil the desire to postulate meta-dynamism as a real physical aspect of the world. Meta-dynamism was not an afterthought to reifying dynamism.

The arguments summarized here are not ones of secure logical deduction. But they do not depart in that respect from most arguments about consciousness, and I claim that they are at least highly suggestive.

2.4 The Necessity Claim

The argument about achieving PRAIS that has just been outlined leads to the *Necessity Claim* of MDyn, which is that

> *a conscious process must, at each moment in its time-span, be meta-dynamically sensitive to (at least recent) dynamism within the process up to the current moment, and in such a way as to enable the process to individuate itself. That is, it must, by* ***meta-dynamic*** *means, be continuously, auto-individuatingly auto-sensitive.*

The following clarifies some aspects of the claim.

- The Necessity Claim only requires auto-sensitivity to "at least recent" dynamism. The intent here is not to demand that it be sensitive to its whole history up to the present, but only perhaps from some little way back up to the present. So the Necessity Claim could be construed as being about a sliding window of consciousness in a conscious process, rather than about the whole process. However, future developments may establish that this sliding window is enough for the process to be able to stitch together a sense of itself as a whole.

- The meta-dynamic auto-sensitivity is itself part of the process's dynamism, and as such can in turn be part of the dynamism that the process is sensitive to. As a result, the theory leads naturally to he idea that the meta-dynamic auto-sensitivity is, at least sometimes and at least in part, constituted of *auto-meta-dynamism*. Auto-meta-dynamism is meta-dynamism that is sensitive to itself (which is the same as to say it affects itself). The sensing/affecting is part of that very meta-dynamism, so the reflexivity in the "auto" can be said to be *intrinsic*.

- The Claim is indeed merely one about necessity, and of course there may be further conditions that a process's meta-dynamic auto-sensitivity might need to satisfy in order for the process to be conscious. Thus, all the Claim says is that *some "suitable" form of* meta-dynamic auto-individuating auto-sensitivity is required. What the suitability amounts to is yet to be determined. It may, perhaps, involve a particular qualitative type of sensitivity, and/or on some particular arrangement of qualitatively different auto-sensitivities, and/or a threshold of intensity.

- The Necessity Claim does not specify precisely what sensitivity or individuation amounts to. It is thus a condition that can be made more precise in different ways by further philosophical development, or physical-theoretical development.

2.5 The Sufficiency Conjecture

The Sufficiency Conjecture is that

> *having some suitable form of continuous, meta-dynamic auto-individuating auto-sensitivity is enough to make a (physically possible) process conscious in at least some minimal, crude, but still phenomenal*

sense. This is not about mere logical sufficiency. Rather, the suitable meta-dynamic auto-sensitivity IS the process's conscious phenomenality.[6]

Some additional comments are as follows.

- The minimal form of consciousness is conjectured to cover certain very basic phenomenal "feels", including a sense of the process's own continuing existence, and perhaps primitive forms of pleasure and pain.
- But for rich, higher-level forms of consciousness it is expected that the process will need to satisfy much stronger conditions than merely to have the suitable meta-dynamic auto-sensitivity.

The Sufficiency tenet is labelled as a Conjecture rather than a Claim because MDyn does not have specific arguments to it from premises, unlike the case of the Necessity Claim.

2.6 Adjuncts to the Sufficiency Conjecture

If one supposes the Sufficiency Conjecture to be true, there are "riders" or "adjuncts" that one can add, giving more specific forms of the theory. They are not just arbitrary, but motivated by the thinking leading to the Necessity Claim. Two such are:-

Adjunct 1: Existence of Core Meta-Dynamism and Consciousness: The suitable, continuous meta-dynamism in the Sufficiency Conjecture has a basic version (or just is a basic version) that is a core, minimal form of consciousness, necessarily present in all consciousness, however primitive.

Adjunct 2: Quality of the Minimal, Core Consciousness

The core consciousness proposed in Adjunct 1 is contentless except for being a crude feel of the experience's own continuing existence (however brief) and perhaps for including basic versions of pleasure and pain.

Adjunct 2 makes a form of temporal consciousness core to all consciousness. This form is basic and pre-reflective. It does not, for instance, involve a concept of time durations, successions, etc. Rather, it is a direct experience of the temporally local durating (continuing).

The point of Adjunct 1 is that the Sufficiency Conjecture in itself does not imply that all conscious processes contain a common form of basic meta-dynamic auto-sensitivity. In a way, Adjunct 1 amplifies the Conjecture to one that could have been presented in the first place. However, it is theoretically useful to separate the issues as above. The adjunct also serves to make the common core be the form of meta-dynamic auto-sensitivity required by the Necessity Claim.

These adjuncts connect to considerable interest in consciousness studies in various forms of core, minimal or pure consciousness (as cited in the Introduction).

2.7 A Note on MDyn's Anti-Humeanism

On the question of anti-Humean stances in general, I have been particularly swayed by the physics-informed considerations of Maudlin [37]. However, he does not go so far as the radical reification of dynamism in MDyn.

It should be noted, though, that this radical reification, while it could be said to reify law-governedness, does not reify physical laws as such. Laws remain as human-crafted descriptions that reflect the objective regularities that dynamism creates, and that could be varied in many detailed ways.

---

[6] So, if the conjecture is true, the theory is an "identity theory" about consciousness, but because of the enormous multiple realizability, it does not make consciousness identical to states of bodies or brains as in typical in identity theories, but, much more broadly, to configurations of meta-dynamism.

3 From the Philosophical to the Physical-Theory Side of MDyn

Before we arrive at the mathematical treatment, we need to consider some simplifying assumptions and some ways of making the above ideas more specific. All the simplifications and specifications are subject to change in the future.

3.1 Strictly and Loosely Conscious Processes

A simplifying assumption is that one can define the notion of a *strictly* conscious process, in that the activity at all spacetime locations involved in the process are part of the constitution of the consciousness. A loosely conscious process is then one that (in a sense defined below) contains a strictly conscious process but is not itself strictly conscious. Thus, it my be that when someone is engaged in a conscious experience, their whole brain could be only loosely conscious, with some of its activity being unconscious. Even more so, a whole person is, presumably, merely loosely conscious. However, the distinction may not be a hard and fast one and may need to be relaxed in future.

The treatment below is almost exclusively about strictly conscious processes. For one thing, it is a heuristic, pragmatic matter whether any given process should be said to be conscious in the loose sense. For instance, a car being driven by a person contains the strictly conscious processes within the people in the car (including hopefully the driver). But although we say each person as a whole is conscious, we don't normally say the car together with its passengers is conscious, unless we are speaking figuratively. But it is hard to see any precise principle of distinction being used here.

The description of strict consciousness above was in terms of on/off consciousness, but works also on the understanding that "consciousness" there means consciousness of any positive degree.

3.2 Processes in General

I will assume that a process consists in part of the values of states over some specific spacetime region. At each point in the region, all aspects of ordinary world state are included, as far as the current treatment goes. However, this is a simplifying assumption—in the future it would be useful to be able to confine a process to certain aspects of the world. One example of this might be if the process involved only, say, electromagnetic fields and the states of charged particles. However, it is unclear whether this is the right sort of limitation, so the issue is left to further research.

Currently I do not extensively address the question of interactions between the process and its immediate environment, i.e., input to and output from the process. I make some comments in passing below. I currently regard the trajectory of total-states that the process goes through as including the process's side of the input/output interactions. The interactions are key to the issue of auto-individuation, and will play a role in the next Section, but that section makes clear that more development is needed on the matter.

Given that the treatment currently includes all aspects of state a given location, input/output is only between the process's states and states outside the region the process occupies. However, if we allowed processes to involve merely some selected aspects of state, as suggested just above, then input/output could also be with respect to the aspects not included, via laws that link aspects included to aspects not included.

3.3 Conventional Spacetime and Dynamism as Field

I do not attempt at the current stage to bring relativistic considerations into play, apart from making a brief suggestion in the Discussion section about the general-relativity aspect of Orch OR theory of consciousness. This limitation is partly for simplicity at the current early stage of development. But it also because at present it is good enough for MDyn, and probably any other theory of consciousness, to focus on conscious processes within which relativistic effects are presumed insignificant. (Again,

this is apart from the Orch OR exception). To be sure, MDyn along with any other physicalist theory of consciousness will eventually need to encompass the question of relativistic invariance. This includes the question of how to ensure that the mathematical conditions for a process to be conscious in one frame also apply when translated into the viewpoint of another frame moving at non-zero, and possibly high, relative speed.

On this basis I assume classical spacetime coordinates $(\mathbf{x}, t)$ relative to some arbitrary origin. I will usually summarize $(\mathbf{x}, t)$ as being a (spacetime) location l.

I assume that dynamism consists of a *dynamism element* $\mathcal{D}(l)$ at each spacetime point l. Again, this pointwise locatedness is a simplifying assumption for the present stage of development, and in the Discussion section I will suggest a departure from it. A dynamism element is something different from any ordinary aspect of state as above. Also, it may have qualitative features rather than purely quantitative ones (but I will assume quantitative functions that work on the qualitative values, when drawing upon a dynamism element in laws). Roughly speaking, the dynamism element consists of the dynamism at l, where this located dynamism can be thought of as the affectedness by states at nearby locations with earlier times. I assume that for a given region consisting of such locations, the value can be "restricted" to apply only to that region, i.e. to be the affectedness by which locations in that specific region are affecting state at l. In future developments it may also analogously be useful to assume dynamism elements include the affectedness of states at nearby locations with later times by state at l.

I also assume that each dynamism element has an overall *dynamism intensity* (notated by κ below).

I view the fact that dynamic elements consist of affectednesses by spatiotemporally (typically nearby) locations make them into entities that are, inherently, *temporally non-local*. One might perhaps be able to take a philosophical view whereby in some sense the nearby locations are replicated inside the dynamism element, but to me this is yet more mysterious than saying that some physical entities are inherently exist over regions and are not definable at points.

Further than this I need to leave the intrinsic nature of dynamism elements as metaphysically mysterious, except to the extent I have discussed it philosophically above in general intuitive terms, and to the extent that the way dynamism is juggled in the theory being presented itself constrains its intrinsic nature. But perhaps its intrinsic nature may ultimately be perceived to be no *more* mysterious that that of say, mass, energy, fields, etc.

3.4 Towards Meta-Dynamic Laws

The mathematical treatment centres on how meta-dynamism might be reflected in laws as couched in mathematical equations, and how the notion of meta-dynamic auto-sensitivity of processes might be accommodated. I emphasize that (with an exception in the Discussion section) I do not propose particular modifications of any particular law. My intent is just to show the general *sort* of modification that I envisage in order to give a possible precise, fundamental basis to MDyn. Least of all do I attempt to specify how laws would work in concert across a sizable system such as a brain.

I will say that mathematically formulated laws concerning change over time, such as $F = ma$ or some version of the Schrödinger equation, *reflect* some aspect of dynamism. Notice that such a law does not *refer to* or *describe* dynamism as such—it is not *about* dynamism as such. Rather, it is about spacetime-located values of force, mass, acceleration, the prevailing wave function, the electric field, curvature, energy density, etc. etc. (including here derivatives of quantities with respect to time or space). I will call such aspects of the world the "ordinary" (or "base-level") aspects of the world. Some of these aspects are, directly, matters of change over time, such as a rate of change of momentum with respect to time. The changes are produced by the dynamism, and so the laws are dynamic in that sense, but the law is merely *about* what the rates of changes are and how they relate to each other and to other quantities, not about the dynamism itself.

Laws reflecting *meta*-dynamism, by contrast, are ones that are explicitly about dynamism itself, and usually about ordinary world state as well. The laws contain mathematical elements of some sort that refer to dynamism. The elements proposed below involve the dynamism elements $\mathcal{D}(l)$ above.

I will say that laws that are (at least partly) about dynamism, and therefore reflect meta-dynamism, are *meta-dynamic* laws. There are two broad possibilities for such a law: (1) it is a directly modified form of some existing law, e.g. a modified classical law, a modified Schrödinger equation in quantum theory, or a modified Einstein field equation in general relativity, with for instance extra terms that refer to dynamism; or (2) it is not such a modification—it is a completely new law. Under (2) we could have laws that provide constraints between ordinary state and dynamism elements, and/or laws that are just about dynamism, without constraining ordinary state. I will be concentrating on (1) in Section 4, while giving some indication of possibilities for (2) in Section 5.3. In (1), I include not only possible changes to existing, standard laws, but also changes to already non-standard laws as proposed by other researchers. These non-standard laws might themselves be variants of standard ones or might be new ones.

Just as base-level, or ordinary, dynamism is what is reflected in existing laws of physics such as the Schrödinger equation (those laws show, so to speak, the shape of changes, produced by ordinary dynamism, of the ordinary quantities the laws are about), meta-dynamism is what is reflected in meta-dynamic laws (they show the shape of changes, produced by meta-dynamism, of the aspects of dynamism that they are about). A complication is that a meta-dynamic law may not be about, or only about, base-level dynamism, and thereby reflect a low-level aspect of meta-dynamism, but they may also be *about* meta-dynamism itself, and therefore reflect a higher level of meta-dynamism. This talk of levels is impressionistic, and below I comment further on the matter. Note that meta-dynamic laws may also reflect base-level dynamism if they are partly about ordinary aspects of the world—as is necessarily the case for laws of type (1).

All this makes for a complex, reflexive view of the lawful nature of the universe, because dynamism is now not just something to which the operation of laws implicitly contributes but also something that can be explicitly manipulated by them. It can even be the case that a law can explicitly operate upon the very type of dynamism that that law's operation implicitly contributes to at the same spacetime location.

I assume that a law that mentions a quantity q, and in fact is sensitive to or affects q's value at a given l, contributes implicitly to the dynamism of the world at l. (Normally, any law that mentions a quantity will help to establish its value at each time, but the treatment below raises the possibility of exceptions.) Similarly, if a law mentions some element of dynamism, it is being implicitly *meta*-dynamic in being sensitive to or affecting that element. But this is different from the *explicit* effect, itself, that the law imposes on that element. So the law is making both an implicit and an explicit contribution to overall dynamism.

3.5 Relationship to Quantum Theory

The treatment below, and the comments above, are mostly expressed in ostensibly classical terms, but I believe it will at some point be able to be absorbed with adaptations into quantum theory, largely by replacing talk of ordinary states by talk of (aspects of) the quantum state as described by the quantum wave function. MDyn is most friendly to quantum theory as equipped with an ontological interpretation. I am particularly influenced by the Bohmian approach here, though nothing in this article depends on the fine detail of that approach. Following [**?**], I take [7] as the main reference point of the approach (partly because of that work's material on consciousness, and its related material on the "order of implicate orders," even though the present article does not touch on those aspects). So, I assume that the equations of quantum theory do in some way describe a reality that evolves in a definuite way as defined by some form of Schrödinger's equation, and are not just a calculation tool. Additionally, measurememt/observation is to be brought into the dynamic theory in an integrated

way as just another part of what is going on overall, rather than being artifically separated out as something that is somehow external to a system being studied. This is claimed to solve weirdness issues concerning the Heisenberg uncertainty principle, superposition, Schrödinger's Cat, etc.

4 Towards Basing MDyn in Physical Laws

4.1 Some Initial Definitions

$\mathcal{O}(l)$ denotes the conglomeration of ordinary state values at a spacetime location l. What values this conglomeration contains is left unspecified here—it can vary across different completed physical theories. It includes all ordinary quantities, including relevant derivatives, that are needed in the laws of the theory.

In line with the comments in Section 3.4, I take $\mathcal{D}(l)$ to be the dynamism element at l. $\kappa(D)$ denotes the *dynamism intensity* of the dynamism element D.

Definition 1. ***TotalState****(l) denotes the combined ordinary/dynamism element conglomeration, the pair* $(\mathcal{O}(l), \mathcal{D}(l))$*. A combined value with this structure, even if it us a hypothetical value, is called a* ***total-state.*** *The* ***actual*** *total-state at l is TotalState*(l)*. Given a total-state* T*,* T_o *is its ordinary-state part and* T_d *is its dynamism-element part.*

I assume that there is a metric M_o giving the distance between any two ordinary-state conglomerations, and a metric M_d measuring the distance between any two dynamism elements. (NB: This metric is probably somehow correlated with the relative dynamism intensities of the two values, but is not here assumed to be based on those intensities, or vice versa. However, I will assume a simple connection for a particular purpose below.)

Definition 2. *The* ***metric*** M *on total-states is* $M(T, U) = M_o(T, U) + M_d(T, U)$.

4.2 Meta-Dynamic Variants of Existing Laws

A schematic illustration of the proposed way of modifying existing laws is as follows. Consider an existing law with the structure:

$$u(l) \quad = \quad v(l)\, w(l)$$

A modified version might then have the structure

$$u(l) \quad = \quad \mathcal{G}_1(\mathcal{O}(l), \mathcal{D}(l))\, v(l)w(l) + \mathcal{G}_2(\mathcal{O}(l), \mathcal{D}(l))$$

(or simpler versions where the $\mathcal{G}_1$ or $\mathcal{G}_2$ element is missing). There are additional, related modification possibilities. For example, some existing term in a law could be raised to a power involving an expression of form $\mathcal{G}_i(\ldots)$, or some new operator could be applied to a term, where the effect of the operator is modulated by some $\mathcal{G}_i(...)$ expression. However, for simplicity of presentation I will confine attention to multiplicative modifications on the pattern of $\mathcal{G}_1(\ldots)$ above and additive ones on the pattern of $\mathcal{G}_2(\ldots)$.

The $\mathcal{G}_i$ expressions written above stand for *gating expressions* in the particular laws. The expressions as shown are just schemata, standing for whatever detailed expressions about ordinary state or dynamism at l are needed in the particular law. The $(O)(...)$ portion of the schematic $\mathcal{G}$ expression

indicates that the actual expression in the law can mention any ordinary state values at l. Similarly, the $\mathcal{D}(l)$ portion indicates that the actual expression in the law can mention any aspect of the dynamism element at l. The basic motivation for including gating expressions is to allow the sensitivity of dynamism (at l) to ordinary state (at l) that is mentioned outside gating expressions in the law, or the sensitivity of the latter to the former, to be controlled or "gated" in whatever way and to whatever extent is appropriate.

I assume that the actual gating expressions in laws are dynamically reasonable in the sense of expressing functions (*gating functions*) that are sufficiently continuous and differentiable with respect to space, time and other values as needed.

Naturally, existing physical laws enable most of the world's changing to be specified to a certain (high) degree of precision. So, meta-dynamic effects should be significantly large only in exceptional circumstances, e.g. only at places and times where exceptional physical circumstances hold. (Within the exceptional cases will be certain brain regions while the person or other creature is conscious.) Hence, over spacetime regions where there are under "normal" circumstances, we would like the varied laws to operate exactly or almost exactly like the original ones. I will continue this topic in Section 5, but for now I note that for a varied law to fulfil that condition, we need the value of an additive gating expression such as the $\mathcal{G}_2$ above to be, over the region in question, zero or a tiny value (scalar or otherwise) relative to other values involved, and the value of a gating expression that is a multiplicative factor such as the $\mathcal{G}_1$ above or a power to be (dimensionless) 1 or a value extremely close to it. I will say that being sufficiently close to these 0 or 1 values makes the gating expression *invisible*, while other values make it *visible*.

If a gating expression value is exactly 0 or 1 as appropriate—i.e., depending on whether the element is additive, or multiplicative, etc.—I will say it is a *absolutely* invisible value, and the gating expression in question is absolutely invisible (at the current place and time). For some purposes, including consciousness theory, it would be useful to know whether and how gating expressions can be invisible in this absolute sense over non-punctate regions of spacetime, as opposed to just having an absolutely invisible value at occasional spacetime points. I will take up this matter in below. Note that even absolute invisibility does not mean the ordinary state and dynamism state mentioned by the gating expression are protected from taking part in interaction with the ordinary quantities elsewhere in the law: those quantities may be viewed as helping to constrain the gating expression to be absolutely invisible and thereby constraining the state items it mentions.

As a special case, it may be that only the $\mathcal{O}$ portion in a gating expression affects whether the element is visible or not—that is, it is only present ordinary state flavours that provide switching between invisibility and visibility. Another special case is that, when a gating expression is visible, its value at a particular l may depend only on the $\mathcal{D}$ portion—i.e., the gating expression is sensitive, when visible, only to the dynamism, not ordinary state.

A combination of the two special cases is that the ordinary state provides all switching and only switching. However, other possibilities may be natural, such as that, when ordinary state is far from being suitable for switching, then the gating expression is invisible, but when ordinary state is near to being suitable, dynamism can also help with switching; and when the element is visible, its value may depend inextricably on both dynamism and ordinary state.

Recalling the point above that, amongst others, Bohm and Hiley [7] suggest the possibility of modifications to quantum theory, it is useful to mention some mathematical formulations from that particular work and how they might be modified in our terms. First, we take their formulation of the equation of motion of a particle of mass m in a classical potential field $V(l)$:

$$m\frac{dv}{dt} = -\nabla(V + Q)$$

where v denotes the particle's momentum. Bohm and Hiley call Q a "quantum potential." It is derived straightforwardly from the Schrödinger equation for a particle (*ibid.*, pp.28–30). Without the Q the equation is equivalent to the classical $F = ma$. But, According to Bohm and Hiley (p.345), one could bring in arbitrary extra forces F to get

$$m\frac{dv}{dt} = -\nabla(V + Q) + F$$

A close parallel using a gating expression would be

$$m\frac{dv}{dt} = -\nabla(V + Q) + \mathcal{G}(\ldots)$$

where now F is mentioned within the gating expression. Or we could add a gating term to the $V + Q$.

Bohm and Hiley [7] also consider a similar manipulation in terms of quantum field theory (*ibid.*, pp.241, 379). If one considers a classical linear field equation

$$\frac{\partial^2\phi}{\partial t^2} = \nabla^2\phi$$

then the quantum theoretic version becomes

$$\frac{\partial^2\phi}{\partial t^2} = \nabla^2\phi - \frac{\delta Q}{\delta\phi}$$

by a treatment directly analogous to that leading to the above equation of motion, and with Q here analogous to the Q above. So we could suggest including analogous gating expressions here.

Section 5.9.2 makes a related suggestion about a gating modification to an already-modified Schrödinger equation, one proposed by Kremnizer and Ranchin [33] in order to provide an objective dynamic model of quantum-wave collapse.

4.3 The Reflexivity of the Meta-Dynamism

The intensely reflexive situation as regards the relationship between laws and dynamism that was mentioned in Section 3.4 is schematically represented by the constraint that the actual total-state at l, i.e. TotalState(l), must be a solution for T in :

Fixed Point Equation constraining Total Dynamism

$$\Lambda(T) = T_d$$

where the symbol Λ is used because it alludes to laws. $\Lambda(T)$ for a total-state T returns the dynamism element describing the dynamism that would have to exist for T to be actual (given the total-states surrounding but earlier than l). The dynamism referred to here incorporates both the explicit effect on

dynamism of all laws via gating expressions and the implicit dynamism inherent in the operating of the laws. So, if T *is* the actual state, it must satisfy $\Lambda(T) = T_d$.

I leave the detailed nature of Λ to future development of the proposal. However, I assume that Λ is continuous with respect to the metric M on total-states and some metric or topology yet to be devised for the space of dynamism elements. Also, some of the following comments constitute further restrictions.

I assume that as an intrinsic effect of dynamism, it leads to a total-state automatically arising at l that satisfies the fixed-point equation and maximizes the continuity with the solutions at nearby spacetime locations that have earlier times. Again, the details of this must await future developments. It may be that we should also mix in an element of minimization of the intensity of explicit dynamism in the total-state.

The developments below, while implicitly informed by the fixed-point equation, makes more explicit use of the following concepts.

Reflexive Constraint on Ordinary Dynamism

For any total-state T:

$$\Lambda_o(T) \subseteq_d \Lambda(T)$$

$$\kappa(\Lambda_o(T)) \leq \kappa(\Lambda(T))$$

.

Λ_o returns a dynamism element that describes only, and all of, the *ordinary-level* (base-level) dynamism present in $\Lambda(T)$, i.e. dynamism that is in no way meta-dynamic. κ is the dynamism-intensity function mentioned above. The relation $\subseteq_d$ says that the dynamism described by the left-hand dynamism element is included in (or: part of) the dynamism described by the right-hand dynamism element. Again, the detailed nature of this relation must be left to future developments. The stated constraint therefore merely codifies this "present in" intuition, plus the important point that we are assuming that there is an *objective, physical difference* between the ordinary-level dynamism and the remaining dynamism, i.e. the meta-dynamism.

The metric M above now enables the following definition:

Definition 3. *The **meta-dynamic intensity** of a total-state T is $\kappa_m(T) = M_d(\Lambda(T), \Lambda_o(T))$.*

If there are no gating expressions in any law, so none are meta-dynamic, then the value of $\Lambda(T)$ must not be affected by T_d and should just consist of the the normal, implicit, ordinary-level dynamism inherent in the running of laws. So the fixed-point equation and the reflexive constraint on ordinary dynamism are satisfied with TotalState$(l)_d$ being just that dynamism. In effect the equation and constraint have no added value, which is clearly a desirable state of affairs for the case at hand.

From our terms "meta-dynamic intensity" and "dynamism intensity" it is reasonable to expect that the natures of the metric M_d and of κ need to be mutually constrained in such a way that $\kappa_m(T) \leq \kappa(T_d)$. We will assume a simple connection in section 4.7 for a special purpose, but that particular connection may be too simple.

For the purposes below we actually need an extended, region-specific version of the above notions. For this we use a notion of a "relative" region, a set of displacements that can be applied to any spacetime point to take us to other (typically nearby) points.

Definition 4. *A **relative temporally-abutting region** is an open set of displacements* $[\Delta \boldsymbol{x}, \Delta t]$ *with* $\Delta \boldsymbol{x} \neq 0$ *and* $\Delta t < 0$*, and furthermore if* $[\Delta \boldsymbol{x}, \Delta t]$ *is in the set then so is any closer displacement (i.e. any with a smaller non-zero* $\Delta \boldsymbol{x}$ *and/or greater non-zero (i.e., closer-to-zero)* Δt*).*

That is, from any given spacetime point, each displacement takes us to some spatially different point at an earlier time; and the points thereby found are a dense open region touching l. For convenience, I use notations like "ΔR" to denote the relative regions.

Then, let $\Lambda(T, \Delta R)$ and $\Lambda_o(T, \Delta R)$ respectively denote the total and ordinary dynamism restricted by the relative temporally-abutting region ΔR, in the sense of restriction that was introduced in Section 3.3. I do not assume that the restriction is defined for all such regions, but defined at least for sufficiently small regions. Then I assume that the region-specific version of the above reflexive constraint holds:

Reflexive Constraint on Ordinary Dynamism: Region-Specific

For any total-state T,

$$\Lambda_o(T, \Delta R) \quad \subseteq_d \quad \Lambda(T, \Delta R)$$

$$\kappa(\Lambda_o(T, \Delta R)) \quad \leq \quad \kappa(\Lambda(T, \Delta R))$$

.

for any relative temporally-abutting region ΔR for which those dynamism elements are defined.

Than we can define also a region-specific version of meta-dynamic intensity:

Definition 5. *Given a total-state* T*, its **meta-dynamic intensity with respect to** a* ΔR *as above is* $\kappa_m(T, \Delta R) = M_d(\Lambda(T, \Delta R), \Lambda_o(T, \Delta R))$*.*

4.4 Towards Consciousness: Processes in General

So far in this section we have largely addressed meta-dynamism in complete generality, not looked at specific requirements for the meta-dynamism that MDyn claims is involved in (strict) consciousness. The primary matter to consider is the Necessity Claim. We will put this claim into the more precise form of the Necessity Condition below.

First, I present the notion of a (not necessarily conscious) process suitable for the present initial treatment of consciousness.

Definition 6. *For a spacetime region* R*,* ***Times****(R)* $= \{t \mid \exists(\boldsymbol{x}, t) \in R\}$*.*

Definition 7. *A non-instantaneous region R of spacetime is **temporally connected** if and only if* $\forall t_1, t_2 \in$ *Times(R)* $\forall t : t_1 < t < t_2 \quad \exists(\boldsymbol{x}, t) \in R$*.*

Definition 8. *A **process** P is a pair (Home(P), ProcStates(P)) where*

Home(P) is a non-instantaneous, temporally connected, open region of spacetime, called the home region of P.

ProcStates(P) is a function from Home(P) into the set of conceivable total-states.

Definition 9. *A process P is **actual** if and only if* $\forall l \in$ *Home(P), ProcStatesP)(l) = TotalState(l).*

The defined notion of a process omits any constraint on the state trajectory (as given by the TotalState function), in order to allow hypothetical consideration of processes that may not be physically possible. Where necessary, an extra assumption can be made about a process under discussion that it is a

physically realizable one (if its spatiotemporal origin is suitably positioned in a physically possible universe), or, for instance, that the trajectory is continuous in space and time.

Definition 10. *For a process P,* ***Times****(P) = Times(Home(P)).*

Definition 11. *A process P is a* ***slimmed version of*** *a process Q if and only if Home(P) ⊆ Home(Q), Times(P)=Times(Q), and ∀l ∈ Home(P) (hence also automatically in Home(Q)), ProcStatesP)(l) = ProcStatesQ)(l).*

Definition 12. *A process is a* ***proper*** *slimmed version of Q if and only if P is a slimmed version of Q but unequal to it.*

Notice that because Home(P) is an open set, for every location in it there is an open ball in spacettime containing that location and lying within Home(P). When we consider locations nearer and nearer to the spatial border of the home region at a given time, the ball will generally need to be smaller in order to stay within the home region. We are not including points on a sharp border in the home region, but, the nearer and nearer we are to the border, the closer we are to "input/output" interactions with the outside environment of the process.

The following definitions are needed for the following subsections.

Definition 13. *Given a process P and $l \in$ Home(P), let an* ***inner abutting relative region*** *of P at l be a relative temporally-abutting region ΔR such that $\forall[\Delta \boldsymbol{x}, \Delta t] \in R$ we have $(\boldsymbol{x} + \Delta \boldsymbol{x}, t + \Delta t) \in$ Home(P). (I.e., the displacements in ΔR keep us within the home region, and go only into the past portion.)*

Definition 14. *Let a* ***relative half-ball*** *be a relative temporally-abutting region such that for some $[\Delta \boldsymbol{x}_0, \Delta t_0]$, the displacements in the region are all the displacements $[\Delta \boldsymbol{x}, \Delta t]$ with $0 < \Delta \boldsymbol{x} < \Delta \boldsymbol{x}_0$ and $0 > \Delta t > \Delta t_0$. The* ***size*** *of the half-ball is $[\Delta \boldsymbol{x}_0, \Delta t_0]$.*

Definition 15. *Given a process P and $l \in$ Home(P), let an* ***inner relative abutting half-ball*** *of P at l be a relative half-ball that is also an inner abutting relative region for P at l.*

4.5 Towards Consciousness: Sensitivity to Dynamism

Now we are ready to address meta-dynamic auto-sensitivity.

Definition 16. *A process P is* ***meta-dynamically auto-sensitive*** *at $l \in$ Home(P) if and only if for all inner relative abutting half-balls ΔH of P at l below some size we have κ_m(ProcStatesP)(l), $\Delta H) > 0$.*

That is, roughly speaking, there is meta-dynamic auto-sensitivity at a place and time in the process if the process continues to have some local meta-dynamic intensity, i.e. meta-dynamism intensity with respect to its own past locations, no matter how close to that place and time we get.

The definition makes no mention of auto-individuation. It is not clear whether auto-sensitivity as defined, or with minor developments, is sufficient for auto-individuation. A brief suggestion on this is made in Section 5.2. For now we can note that the definition may supply something that is helpful for auto-individuation. By the above comment about points nearer and nearer to somewhere on the spacetime border of P, any meta-dynamism inherent in the input/output does not account for all the meta-dynamism. (So nor does any dynamism strictly outside the process.) This is under the assumption that any dynamism in which the process should be sensing in auto-sensitivity is separate from the dynamism involved in input/output.

However, the above definition allows the local meta-dynamic intensity to get arbitrarily small as we approach l. There would seem to be more justification for saying that input/output did not exhaust the meta-dynamism at l if there were a positive lower bound to the κ_m values, no matter how small the half-balls get. (This lower bound may depend on the particular process P.) I incorporate this idea in the following definition.

Definition 17. *A process P is **boundedly meta-dynamically auto-sensitive** at $l \in$ Home(P) if and only if $\exists \mu > 0$ such that for all inner relative abutting half-balls ΔH of P at l below some size we have κ_m(ProcStatesP)(l), $\Delta H) > \mu$. (NB: μ can depend on the particular l.)*

Definition 18. *When there is such a lower bound at a location l in a process P, we can use the greatest such bound (i.e., the supremum of the possible lower bounds) to be a measure **Auto**(P, l) of **the intensity of the process's meta-dynamic auto-sensitivity at** l.*

(In the future it may be useful to investigate whether the $\kappa_m(\ldots)$ values tend to a positive limit rather than merely being positively bounded below.)

A remaining issue is that μ may get arbitrarily small within the process's home region, so that overall the boundedness does not provide as much separation from the environment as we would like. This motivates the postulation of a location-independent (but still process-dependent) bound in the following definition.

Definition 19. *For a process P, P is **uniformly-boundedly meta-dynamically auto-sensitive throughout** if and only if P is boundedly meta-dynamically auto-sensitive at every l in Home(P) and there is a lower bound $\mu > 0$ that provides the boundedness of meta-dynamic auto-sensitivity at each such location.*

We can now give a simple statement of a precise necessary condition that at least partly fills the role of the Necessity Claim in Section 2.2.

A Necessary Condition for Strict Consciousness:

For a (physically realizable) process P to be strictly conscious (whether in an on/off sense, or with degree greater than zero if consciousness is purely graded), P must be uniformly-boundedly meta-dynamically auto-sensitive throughout.

As indicated above, this may not supply the resources needed for auto-individuation, so it may be conceptually and crucially weaker than the philosophical Necessity Claim in Section 2.4.

On the other hand, one may wonder whether the Condition is too demanding in a certain respect. This respect is that, at every time $t \in$ Times(P), it requires meta-dynamic auto-sensitivity at *all* spatial places involved at t in the process. It might be thought that one could have a strictly conscious process where at each time only some spatial places are responsible for the auto-sensitivity. But there is an argument against this. Suppose for some time t there are values $\mathbf{x}$ for which there is no meta-dynamic auto-sensitivity at $l = (\mathbf{x}, t)$. Then, unless l is crucially involved in a yet-to-be-discovered necessary condition for strict consciousness in the *remainder* of P, we can remove it from the home region. We remove all such points. Noting the requirements above for the home region of a process, the resulting region is still non-instantaneous and temporally connected (because t is still in Times(P) as we have not removed all points that have time t), so we have obtained a process that is a slimmed version of P that is boundedly auto-sensitive at all places at t. This reduction does not destroy the unfolding of the process, because now we simply regard the interaction with the excluded points as being additional input and output that the slimmed process is engaged in. In sum, as far as can be seen at present, we can safely slim down a process that is not everywhere auto-sensitive at each time to one that is.

4.6 Transverse Continuity of Consciousness

It is reasonable to expect that sufficiently minor perturbations to a strictly conscious process will leave one with a strictly conscious process. More concretely, for a conscious process in a brain, it would be surprising if it were not the case that sufficiently minor perturbations to the states of the neurons, the lengths of axons, the timings of transmissions of signals, etc. etc. leaves one with a conscious process. Thus, we expect some sort of *continuity* with respect to perturbations of state, shape, etc. And this is not something to expect just in the context of MDyn's claims concerning meta-dynamism, but is something that is reasonable for *any* physicalist theory of consciousness to uphold.

But, in the particular context of MDyn, continuity in that sense has an interesting interaction with the Necessary Condition above. For ease of reference, I call that sense *transverse continuity* as it concerns smoothness of behaviour of consciousness with respect to changes in ordinary state and dynamism at each spacetime location l, and so is about a (non-spatial) "cross-section" of the process at each location. A complication is that transverse continuity allows for smooth perturbations of the spatial and temporal distances between locations within the process's home region, as long as the total ordering of locations along every axis (spatial and temporal) is preserved, e.g. if one location is later than the other then the same applies to their perturbed versions.

The following postulate about, and incorporated definition of, transverse continuity takes consciousness to be graded, purely or impurely, by a scalar measure $C(P) \geq 0$. However, the definition accommodates purely on/off consciousness if we identify "off" with degree 0 and "on" with 1, and take the ϵ in the postulate to be any value between 0 and 1. I use e is the Euclidian distance metric on spacetime locations. For a function σ on a set S of locations, I use $\sigma(S)$ to mean $\{\sigma(l) \,|\, l \in S\}$.

Postulate of Transverse Continuity of Strict Consciousness: Graded Case

The following applies to every strictly conscious process P (i.e, where $C(P) > 0$) and every physically possible process Q.

$\forall \epsilon > 0 \;\; \exists \delta_1 > 0, \delta_2 > 0$:

IF there is a continuous, order-preserving map σ on the spacetime locations in P such that Home(Q)=σ(Home(P))
and $\forall l \in$ Home(P) we have $e(\sigma(l), l) < \delta_1$ and M(TotalState(P)(l), TotalState(Q)($\sigma(l)$)) $< \delta_2$,

THEN we also have $|C(Q) - C(P)| < \epsilon$
and so Q is strictly conscious if $C(P) - \epsilon >$ the lower bound for C (which is 0 when consciousness is purely graded).

The defined continuity is non-uniform with respect to processes P, in that how small δ_1 and δ_2 need to be can depend on which P it is. However, for a given P, the continuity is *uniform with respect to the different locations* l in P's home region, in that given δ_1 and δ_2 values need to apply across all the locations. I discern no reason at present for suspecting that this uniformity is too strong a requirement or that the non-uniformity with respect to P is too weak.

The uniformity of transverse continuity with respect to locations creates the following interaction between the continuity and the necessity of meta-dynamic auto-sensitivity in consciousness.

4.7 Interaction between Transverse Continuity and the Necessary Condition

The intuition about the interaction is that we might have a strictly conscious P and a Q with the same home region that lacks P's meta-dynamic auto-sensitivity somewhere in its time-span, but is nevertheless still close enough to P overall that, by transverse continuity, it must also be strictly conscious. This would violate the Necessity Condition as applied to Q.

The conjecture that arises from this is that the Necessity Condition and Transverse Continuity together imply that for any strictly conscious process there is some positive lower bound that the

intensities of meta-dynamic auto-sensitivity at all the locations in the process must achieve or exceed. That lower bound is closely related to how big the δ_2 values can get in the definition of continuity. Notice carefully that, although the notion of boundedly meta-dynamically auto-sensitivity above involves a lower bound on the auto-sensitivity, that bound was not assumed uniform across the locations in the process, so it could get arbitrarily small—it does not imply a positive lower bound that would work across all the locations.

The argument I will give is not precise, but nevertheless suggests that with sufficient extra, reasonable restrictions on how dynamism works, a precise version of the argument would go through. Also, the argument requires the following simplifying assumption and postulate, which appeal to the the dynamism intensity measure κ defined above for dynamism elements.

Simplifying Assumption about M_d: For any dynamism elements D, E, $M_d(D, E) = |\kappa(D) - \kappa(E)|$.

Postulate on Meta-Dynamic Auto-Sensitivity and Dynamic Intensity

For any actual process P and $l \in \text{Home}(P)$, $\quad \text{Auto}(P,l) \leq \kappa(\text{ProcStates}P)(l)_d)$.

The basic reason for taking this postulate to be reasonable is that Auto is measuring the M_d-distance between two parts of P's total dynamism at l, where one part includes the other, and now such distances are defined in terms of κ.

In more detail, notice that Auto is defined indirectly on the basis of distances given by the M_d metric between part of the process's total-state's dynamism element at the location at hand, further restricted by just considering the dynamism at temporally-abutting regions staying within the process, and the whole of the dynamism in the total state. Also note that since P is actual, its total-state obeys the fixed-point equation. So, the in the above use of $\kappa_m(T, \Delta R) = M_d(\Lambda(T, \Delta R), \Lambda_o(T, \Delta R))$ to define region-specific meta-dynamic intensity (Definition 5), where we now specify T to be P's total-state, it is reasonable to expect that this intensity is less than the M_d-distance what we get without the region-specificity, i.e. $M_d(\Lambda(T), \Lambda_o(T))$. But this equals the M-distance between T_d and its ordinary-dynamism aspect, by the fixed-point equation and the nature of Λ_o. But since, by the Simplifying Assumption about M_d, the fact that T_d includes its ordinary aspect, that difference is $\kappa(T_d)$ minus κ applied to that ordinary aspect, so it must be less than $\kappa(T_d)$ itself. But this is just the right-hand side of the above postulate.

Next we need the following definition, of the "tolerance" of conscious processes to perturbation of state values as encapsulated in transverse continuity. This Tolerance is definable because of the uniformity across locations that the above continuity definition possesses. For simplicity I will ignore the possibility of deformation of the region by a function σ. (Its effect can be straightforwardly introduced.)

Definition 20 (When consciousness is at least partly an on-off matter)**.** *For a strictly conscious process P,* ***the tolerance Tol****(P)* = $\sup(\{\delta_2 | \delta_2$ *works as in the definition of on/off transverse continuity to ensure that Q is strictly conscious* $\})$.

Definition 21 (When consciousness is at least partly a graded matter.)**.** *For a strictly conscious process P with positive consciousness degree $C(P)$, and any $\epsilon > 0$,* ***the tolerance Tol****(P,ϵ)* = $\sup(\{\delta_2 | \delta_2$ *works as in the definition of graded transverse continuity to ensure that $C(Q)$ is within ϵ of $C(P)$* $\})$.

Thus, we have that, in the on/off case, if the departure as regards state values of process Q from P is less than Tol(P) then Q will be strictly conscious (because that departure must be less than some δ_2 that works), and if the departure is greater than or equal to Tol(P) then there is no guarantee that Q is strictly conscious (though it may be). Similarly, in the graded case, Q's state-wise departure from P

being less than (respectively, at least as much as) Tol(P,ϵ) provides a guarantee (respectively, provides no guarantee) that Q is within ϵ of P's consciousness degree.

To see the interaction of continuity with the Necessary Condition above, suppose first that strict consciousness is purely on/off. I will continue to ignore the possibility of the spatiotemporal deformations by σ functions.

Consider a strictly conscious, actual process P, and any given location l in its home region. The process is uniformly boundedly meta-dynamically auto-sensitive at that location, by the Necessity Condition, with auto-sensitivity intensity Auto(P, l). Suppose now that Auto(P, l) $<$ Tol(P).

Consider now a slightly perturbed variant Q of P, where the perturbation is that Q has no meta-dynamic auto-sensitivity. Assume that Q could have been actual should the prior history of the world have been slightly different. In effect, we can take Q to be actual(as well as physically possible).

Now, we can realistically expect from this that, at l, Q must depart from P in other respects as well, and that, near l, Q must depart a little from P in various respects. However, let us first suppose as an approximation that Q's only departure from P is precisely the stated departure on meta-dynamic auto-sensitivity, and all the rest of the dynamism at l stays the same. I will assume from this the κ value is reduced just by Auto(P, l). (This reasoning is only valid if Q can be considered actual, and of course would not makes sense at all without the Postulate on Meta-Dynamic Auto-Sensitivity and Dynamic Intensity.)

As a result, and using the Simplifying Assumption about M_d, the M_d-distance between the dynamism elements for Q and P at l is Auto(P, l). As a direct result of this and of Q not departing from P on ordinary values, M(TotalState(Q)(l), TotalState(P)(l)) = Auto(P, l).

But Auto(P, l) was supposed less than Tol(P). Thus, since Q is assumed physically possible and is identical to P in all other respects, Q satisfies the nearness condition in the postulate of transverse continuity, and must therefore be strictly conscious. But of course this violates the Necessary Condition, because Q has no auto-sensitive meta-dynamism at l.

So, this simplified, approximate argument suggests that at each location l throughout P, the intensity of the meta-dynamic auto-sensitivity must be at least Tol(P). This is not dependent on l. So we have argued that, for a given strictly conscious P, there is a uniform positive lower bound, namely Tol(P), throughout P's home region as to the intensity of the auto-sensitivity.

It is easy to see that a similar argument applies when consciousness is purely graded. We now consider a strictly conscious P with consciousness degree $C(P) > 0$, and again a Q that has no meta-dynamic auto-sensitivity at l. Setting ϵ in the continuity definition equal to $C(P)$, we get a contradiction with the Necessary Condition unless throughout P the amount of fully auto-sensitive meta-dynamism is at least Tol(P,$C(P)$). (The impure-grading case is similar.)

I now conjecture that, if we undo the approximative assumptions made about Q's departures from P, we can still expect there to be some positive lower bound, though it may have to be different from the mentioned Tol values. This expectation arises because, for a sufficiently small departure μ of Q from P on the auto-sensitive meta-dynamism at l—i.e., for sufficiently small intensity ν of such meta-dynamism at l in P, given that Q has none such—the departures from P in other respects at l can be expected to be small, and by continuity of state change within P the departures close to l to be small. Also, it would be reasonable to expect the departures to die away to insignificance not far from l. Thus, it would still be the case that Q satisfies the nearness condition in the transverse-continuity postulate. Turning this around, it is reasonable to expect that we still have some positive lower bound ν on the meta-dynamic auto-sensitivity throughout P.

I reiterate that these arguments are merely approximate and suggestive, and rely on special assumptions. But they provide a useful basis for further coherent development of the approach.

5 Discussion and Further Research

5.1 Temporal Non-Locality and Non-Instantaneity

Dynamic elements are time-spanning in their nature, giving the overall proposal a temporally non-local aspect. However, the rest of the proposal has been based on point-locations in spacetime. Because of this tension it may be advisable to make a major shift in the ontology and in the mathematical representation, and abandon point-locations entirely, using only non-punctate regions. Points would become mere approximative and heuristic abstractions, useful for some but only some purposes. This may not only serve the notion of dynamism better, but also fits with Bohm and Hiley's observation in [7](p.374ff) that regions may better serve quantum theory, and with use of regions in work on quantum gravity (see, e.g., [9,53]). The shift would also fit with the more general idea that temporal non-locality in physics could profitably be given more attention [1]. Following [1] one of the possibilities would be a roughly Lagrangian one, involving optimization of suitable quantities over regions. This could serve the need for continuous state trajectories to arise that comply with the Fixed-Point Equation (Section 4.3) everywhere in a region.

5.2 Auto-Individuation and the Necessity Condition

It was noted in Section 4.5 that the mathematical Necessity Condition may fall short of the philosophical Necessity Claim in respect of providing the resources for auto-individuation. A point for further investigation would be to consider whether it would be enough to place a limit on the degree to which the process is sensitive to dynamism that lies across the border with its environment and to dynamism beyond itself.

5.3 Further Sorts of Meta-Dynamic Law

Meta-dynamic variants of standard laws are only one type of meta-dynamic law that one might consider. I mention some considerations to guide future research.

Another conceivable option is a law that, temporarily, directly prevents the presence of a certain type of dynamism (probably over a region of spacetime, not at a point). This action would override the action of other laws that affect that type. This idea potentially raises its own consistency and paradox problems if there is more than one such law, or there could be conditions under which the law would apply to the type of dynamism it itself implicitly supplies. But I am currently not aware of any need (from MDyn or other motivations) to propose such override laws.

Even less so do I consider laws whose operation affects laws as such: whether to destroy a law, construct a new law, or modify an existing law structurally. In any case, I reiterate that I do not physically reify laws, and if one wanted something like such effects, they would have to be obtained in a physically real away that could be *described* as the destruction/creation/modification of laws.

A more reasonable option is laws that are constructed entirely out of gating expressions, or something akin to them. In the extreme, the $\mathcal{O}$ portions of the gating expressions might be vacuous—not actually mention ordinary state at all; and there might be no ordinary physical quantities mentioned elsewhere in the law. These laws would then supply an entirely independent web of constraint on dynamism.

5.4 Levels and Ladders of Meta-Dynamism

There is a sense in which dynamism affects itself through the gating expressions in meta-dynamic laws. The dynamism up to some time, at/around a particular spatial location, is sampled by a gating expression and is in mutual constraint with other entities that the law is about. Of course, unless we want to allow some form of retrocausality, that sampled dynamism can only be affected at the

current instant at which the law is being applied. (But if dynamic elements include future-facing aspects as suggested in Section 3.3 then such aspects can be comprehensively affected.) The running of the law itself constitutes some of the dynamism at/around the current instant. By affecting that running, the dynamism-up-to-now is affecting dynamism. So, in a broad sense, we have what can be called auto-meta-dynamism. Note that that auto-meta-dynamism is itself part of the dynamism we are talking about, on both sides of the statement that "dynamism is affecting dynamism". So one could say there is also, automatically, auto-meta-meta-dynamism. And so on up an infinite ladder of levels of meta-dynamism.

However, this is just a way of thinking about the dynamism that may be convenient for some purposes. In reality, I propose, there is just a holistic meta-dynamism (when there is any meta-dynamism at all) that is sensitive to and affects itself. There is no physically-real infinite regress or ladder, or even an objective division into different discrete "heights" of meta (number of metas in "meta-meta-meta...").

Nevertheless, it may be heuristically useful in some practical contexts for a theoretician, experimenter, philosopher, etc. to focus merely on *significant* degrees of meta-dynamism, whatever it is that significance amounts to in the given practical context. Then, depending on the particular details of the gating expressions, it could be, for instance, that ordinary-level dynamism has a significant meta-dynamic effect on other ordinary-level dynamism, but that that the meta-dynamism constituted by that affecting is only, quantitatively, a small portion of the overall dynamism. And meta-dynamism that is influenced by or influences *that* meta-dynamism may be yet weaker. At some point in such a progression, effects may become to weak to count as significant, or even be entirely eliminated. So, in such situations we may be able to discern discrete heights of meta, and the ladder may have finite height.

5.5 On/Offness and Simplicity

An issue for consideration in depth elsewhere is whether the fundamental nature of reality allows there to be be extended regions in the universe where there is absolutely no consciousness, not even the tiniest amount, throughout the region, even though there are neighbouring regions where there is some consciousness. For instance, can consciousness be completely absent outside brains, though present inside some at some times? In the ON/OFF case, can consciousness ever be OFF over a region of non-zero spatial and temporal extent but ON in an immediately neighbouring region; and in the purely graded case, can it ever be at constant, *absolutely* zero degree over a region, but non-zero, or smoothly rising from zero, in a neighbouring region? Of course, these questions engage with the important philosophical issues of whether some form of panpsychism is correct, i.e. basically whether consciousness, to *some* degree at least, is ubiquitous in the universe, and just having particularly marked forms in, for instance, human brains and possibly within other life forms.

This depends in part on deep (and in some cases controversial) questions such as whether spacetime itself is discrete or continuous, whether the universe (in its spacetime and dynamics) is largely continuous but has some discontinuities (with respect to space or time), and whether, to the extent it is continuous, what particular class of mathematical functions describes it, and in particular whether real-valued non-analytic functions are permitted (i.e., real-valued functions that are not everywhere expressible as a power series in a certain way, in fact as a Taylor series [44]). Bound up with these issues is whether discontinuities and non-analytic features appearing in current physical theories (e.g., to describe phase transitions or quantum collapse events) are merely a symptom of the models being just useful calculation tools to make good-enough predictions about reality, or are intended to be approximate descriptions of reality itself, or are actually intended—ultimately, when completed—to describe reality accurately.

I will content myself here with a simple illustration of the issues. It is closely connected to the issue appearing in Section 4 as regards whether there is a positive lower bound on the intensity of

meta-dynamic auto-sensitivity throughout a conscious process. In the following nothing hangs on the particular definition we gave of the intensity of meta-dynamic auto-sensitivity at a given location (see Definition 18 of the Auto measure), so here I will just assume there is some measure q of a process's overall amount/intensity of such sensitivity.

In Section 4, the question was whether we could infer such a bound from other considerations. But one might propose a version of the theory where a lower bound is just postulated. For instance, one might postulate that consciousness is purely graded and that a process is conscious (in the sense of having a positive degree of consciousness) when q meets or exceeds a threshold q_0, with the degree of consciousness being (say) proportional to $q - q_0$ after that. Clearly, this allows consciousness to be zero throughout extended regions, where q is below threshold, but to vary in neighbouring regions. But there is a discontinuity in the rate of change of consciousness degree with respect to q and this can give rise to a similar discontinuity in the rate of change with respect to spatiotemporal location (e.g., a discontinuity at the region border just alluded to). The question is whether such discontinuities should be considered to exist in actual physical reality.

A variant of the question is that we could use a soft thresholding function that is perfectly smooth (infinitely differentiable everywhere in its domain) but nevertheless, through being non-analytic, still allows consciousness degree to be zero wherever $q \leq q_0$ but to start varying as soon as q_0 is reached. (E.g., taking $r = q - q_0$, the function $f(r) = 0$) for $r \leq 0$ but $= e^{-\frac{1}{r}}$ for $r > 0$ is like this. It is smooth throughout, but fails analyticity at $r = 0$, because all derivatives are zero there, meaning that the Taylor series expansion around that point would deliver zero for $r > 0$ as well as $r \leq 0$). However, exploiting this sort of possibility requires non-analytic functions to be usable in descriptions of actual reality, not just in purely calculational or approximative models. In particular, one might be philosophically troubled by the discontinuity in the *conception* of the function $f(r)$ that happens at $r = 0$.

Going in the opposite direction from smoothness, worries about discontinuity etc. go away if space and time are themselves discontinuous. See, for example, the discussion of discreteness in [53], especially with respect to loop quantum gravity (see, e.g., [9]) and the causal set approach (see [52] for review). It is made clear in [53] that discreteness of volume and area operators in loop quantum gravity does not guarantee that spacetime itself is discrete, but the discreteness of such operators may be enough for the purposes of the present discussion.

Although we have raised an MDyn-specific version of the issues, the issues face most theories of consciousness in some form or another. Certainly, any purely physicalist theory must ultimately be consistent with the types of functions that are countenanced by the underlying assumed physical theory of the universe. But even a dualist approach where there are systematic laws connecting the physical realm and a non-physical, phenomenal realm may well face the problem. Just because the latter is deemed non-physical does not mean it is not framable in a systematic way, even a quantitative way. Then there could be continuity (etc.) requirements on the relationships reflected by those bridging laws. Nor need such requirements be quantitatively framed on the phenomenal side—continuity in general just relies on the relevant domains having a suitable analytical topology of open sets, and this can be purely qualitative.

Putting consciousness aside, similar issues arise concerning whether meta-dynamism in general is present almost everywhere or not, whether it can have discontinuous jumps, etc. But the ubiquity question is much less pressing than in the consciousness case. There seems to be something conceptually remarkable about saying that there is at least an extremely tiny amount of consciousness everywhere in spacetime. By contrast, once one has accepted that there may be some sort of meta-dynamism somewhere, there is no particular *conceptual* bar to thinking it might be everywhere. Of course, there are empirical considerations. There would be ever so slight, virtually ubiquitous, deviations of predictions of behaviour from current physical predictions. This consequence appears to be of a piece with the implications of other proposals, for instance, in the potential modifications to quantum theory entertained by Bohm and Hiley [7] as mentioned in Section 4.2. And, in any case, modifications to physics that involve literally ubiquitous effects, even if they are for normal purposes only extremely

tiny, may be needed to solve open problems such as dark energy or to reconcile quantum theory and general relativity.

Thus, it could well be that we need only worry about ubiquitous meta-dynamism when we are considering other matters that are theorized to connect to meta-dynamism, such as consciousness in MDyn. But even when considering MDyn, ubiquity of meta-dynamism in general would not necessarily be a problem, given that consciousness relies on some very special form of meta-dynamism, which may only arise in very special physical circumstances. Recall in particular that the process needs not just to be meta-dynamically auto-sensitive, but, more particularly, auto-individuatingly so. In addition, there may be necessary conditions for consciousness that have nothing to do with meta-dynamism.

Having said al this, MDyn does potentially amplify the issues in contrast to some other types of theory. Theories that demand complex physical structures or complex information-processing states and manipulations, and theories that do not dissolve quantum wave collapse into general dynamics, may have less of a problem with the issues in this subsection, even if the difficulties do not go away entirely.

Indeed, it is possible, especially if Mdyn's Sufficiency Conjecture is correct, that consciousness is fundamentally something quite simple, relying on simple structures of meta-dynamism, even if they are arise only in special physical circumstances. Certainly, consciousness is allied with extremely complex information processing and neural network structures in humans, thereby enriching the consciousness, but that is just a special, "optional" case. The simpler the nature of consciousness at base, the more fundamental are questions like the spatial one of how widespread in the universe it is and the temporal one of how easy it is to arise in evolution.

5.6 Phased Sequences

Importantly, there can be a sequence of phases of a process with different involvement of meta-dynamism. For instance, it could be that at times up to (or very close up to) some time t_0 there is no or very little meta-dynamism at a particular spatial location (or in a particular region). But then suddenly, or over a relatively short space of time, meta-dynamism could become switched on because gating expressions in laws attain visible values as result of newly arising values for ordinary quantities, say. Conversely, there can be sustained significant meta-dynamism throughout some period, because at each time the gating expressions work similarly, but then there is a switching off. Thus there could be a sequence of phases in which meta-dynamism exists (or not) at a place, with episodes of switching in between that are of short duration relative to the phases.

In particular, a process could switch in and out of consciousness this way, or have phases with radically different levels of consciousness.

Going back to ladders of meta-levels of significant meta-dynamism, the picture above of a sequence of phases with different meta-dynamic characteristics can be enriched, because now we can talk of different heights of significant meta-dynamism in different phases.

5.7 Auto-Sustaining Meta-Dynamism

The points just made lead to the more general observation that, once a gating expression attains a significantly high value after having an invisible one (perhaps an absolutely invisible one), it can enable itself to stay that way even if the triggering conditions cease to hold. A particular consequence of this fact could as follows. If a gating expression obtains a significant value, perhaps just because of normal, non-meta-dynamic changes in the world, and thereby allows new meta-dynamism to arise at t (or a time derivative of an aspect of meta-dynamism to get a new value at t) then, assuming suitable continuity of meta-dynamism with respect to time in most circumstances, there will be some non-zero meta-dynamism for some period from t. This meta-dynamism could then be directly picked up by gating expressions in law applications in times later than t, enabling those gating expressions to avoid

being absolutely or non-absolutely) invisible even though they would otherwise have been so. Of course, conditions could change such as to counteract this auto-sustenance, switching the current phase of meta-dynamism off.

Note that, once the phase is under way, a gating expression can in effect detect meta-dynamism as opposed to ordinary dynamism without explicitly distinguishing between those levels of dynamism. It could be simply be that without the meta-dynamism the overall level of dynamism would have been too low to have a significant effect.

These comments describe a direct sort of auto-sustaining of meta-dynamism in that it is achieved by the process detecting meta-dynamism directly (albeit in the implicit way just mentioned) rather than through distinctly different quantities such as ordinary ones. But, more generally, meta-dynamism could be indirectly self-sustaining via interactions involving ordinary quantities.

Direct auto-sustenance of meta-dynamism has an important role in the following.

5.8 The Additional Conjectures and Minimal, Core Consciousness

In the the description of the philosophical side of MDyn, a sufficiency conjecture was proposed (Section 2.5), and also some further conjectures as adjuncts. The sufficiency conjecture was that "suitable" meta-dynamism within a (physically possible) process's internal dynamism is sufficient for that process to be [strictly] conscious, at least in some minimal sense of consciousness. One further conjecture was to the effect that this minimal type is a core of all consciousness, is identical to a core form of "suitable" meta-dynamism, and is (mainly) a crude feel of the experience's own continuing existence (however brief). The other further conjecture was to the effect that actually the meta-dynamism constituting consciousness is some form of auto-sensitive meta-dynamism, hence a special form of auto-meta-dynamism, i.e. dynamism that senses and affects itself); where indeed the full auto-sensing and the affecting of itself by itself is, precisely, itself or at least part of itself.

Now, when the philosophical Necessity Claim of MDyn (see Section 2.4) was devised and argued, there was no reasoned argument put forward that the required meta-dynamism was of that intensely reflexive sort. Nevertheless, the attempt to couch MDyn in a mathematical way led naturally to the meta-dynamism being of that sort (see Section 3.4). Partly this is because, if some aspect of reality is mentioned in a physical law formulated in a familiar mathematical way, it stands to be both affectable by other quantities and the law and capable of affecting them. Obviously, $F = ma$ does not just constrain a on the basis of a given F and m, it also similarly constrains F and m. Then, because the web of constraints based on laws is inherent in the above fixed-point equation, the mutuality of constraint translates into dynamic mutually of affecting, and hence as a special case to the above intense auto-meta-dynamism. Thus, the second further conjecture is not arbitrary but arises in a natural and principled way as a primary possibility.

We also saw that auto-*sustenance* of meta-dynamism is a natural possibility. This applies in particular to the auto-sensitive meta-dynamism involved in consciousness. We can even, in principle, imagine a situation in which some configuration of ordinary state (and possibly dynamism) kicks off a (strictly) conscious process, whose own dynamism in time ceases to need any support from anything outside the process. Even further, the only thing needed for the sustenance might be the process's own dynamism (not ordinary state), and yet further than that, the auto-sustenance might need only the process's auto-sensitive meta-dynamism. Thus, in principle we could end up with a conscious process *which consists purely of auto-sensitive meta-dynamism*, a form of dynamism that affects and is sensitive only to itself, and does not involve ordinary state or even ordinary-level dynamism. This does not mean it does not need some sort of extra structure to be "suitable" for consciousness. It may be that not all occasions of pure auto-sensitive meta-dynamism have anything to do with consciousness. But some form of the envisaged auto-sensitive meta-dynamism could be a candidate for forms of pure, largely content-less consciousness as discussed by some consciousness researchers.

If there can be such a process of absolutely pure consciousness consisting of that form of pure auto-sensitive meta-dynamism, I would claim that its experienced quality would, at least mainly, be the feel of its own continuing existence as an experience. (As an exception, the extra structure just mentioned might lead to some sort of experienced structure in the consciousness. There is also reason to think that there could be a basic form of pleasure or pain.) There is little to the process other than its own continuation of existence as an experience. This sort of consciousness is plausibly at the core of all consciousness, if we are careful to note that the "own" in "own existence" does not (necessarily) allude to any higher-level, complex, personal self, but just to the very experience being discussed.

However, suppose we wish to allow for the possibility of the process stopping sometime. Then, unless this could happen through additional effects internal to the process, we would need to allow some interaction with the world outside itself (not necessarily spatially outside, if we refine our treatment to allow processes to be concerned with only certain aspects of state at any location). Thus some "impurity" in the form of extra interaction, including with ordinary state, would be involved in the process.

But, in any case, it is perhaps unrealistic to suppose in the first place that the involvement of ordinary state that was needed within the triggering of the conscious episode could *entirely* disappear, because of considerations of continuity. So realistically we would expect that some degree of such involvement will continue throughout the consciousness, mixed in with the auto-sensitive meta-dynamism.

Note, however, that this impurity does not necessarily mean that the experienced quality of the consciousness is affected. Not all changes in the details of meta-dynamism need affect the quality of consciousness: particular qualities may map to non-punctate regions of meta-dynamism space. Notice that I am not here slipping into a dualist view; rather, it could be that consciousness of a particular quality *is* any member of a particular class of arrangements of meta-dynamism, rather than being one specific arrangement. Of course, with more and more outside interaction and more and more involvement of ordinary state, we would expect the quality of the experience to become changed and enriched.

5.9 Interactions with Proposals by Others

Here I briefly consider how MDyn could be enriched by ideas from proposals by others or, conversely, how ideas concerning meta-dynamism from MDyn could contribute to the further development of others' proposals. I focus on the Integrated Information Theory of consciousness (IIT) [41,42], Kremnizer & Ranchin's proposal [33] concerning objective quantum collapse guided by "quantum integrated information" (which is consciousness-related but independently interesting), and the Orch OR (Orchestrated Objective Reduction) theory of consciousness [26].

5.9.1 Cross-Fertilization with Integrated Information Theory

IIT has it that a system's degree of consciousness is given by its "Φ" value. This value is a measure of the extent to which the causal interactions over time within the system are integrated as opposed being a matter of subsystems working more independently. The precise details are not needed for the present comments. IIT shares with MDyn a foundation in the structure of causation within a system, although causation means different things in the two approaches. Since it does not matter to either theory what sort of physical system the causation is within (biological circuitry, computer circuitry, or even just fields in space), IIT shares with MDyn the exceptionally radical multiple realizability of consciousness noted in Section 1. In fact IIT is even more liberal than MDyn, since MDyn firmly places its causation at the fundamental physical level of the world, whereas causation in IIT can be defined at any level of description of systems and typically does not explicitly involve our low-level dynamism at all.

"Causation" in the context of IIT is a matter of conditional probabilities of sequences of states, where a state is ultimately a matter of ordinary physical quantities. So ITT, unlike MDyn, has no reliance on MDyn's claim that there must be more to consciousness than the trajectories of (ultimately) ordinary physical states within the system. Because IIT does not require a conscious system to involve any meta-dynamism as defined by MDyn, IIT cannot be a correct theory of consciousness if MDyn is at least roughly on the right lines. Φ might well measure something important to do with consciousness, such as how richly unified it is, but no level of Φ, no matter how high, could itself guarantee any degree of consciousness at all. But it is still conceivable that, if MDyn is true, some level of Φ is a *necessary* condition for consciousness. Thus, it would be possible to propose that MDyn needs to be enriched by adding this necessary condition to MDyn's existing Necessity condition concerning meta-dynamism, and correspondingly adding a condition concerning Φ into MDyn's Sufficiency conjecture, as well as using Φ to measure something about the strength or richness of the consciousness episode. It is even conceivable that the as-yet-unresolved question about what particular sort of meta-dynamic auto-sensitivity MDyn needs (the to-be-determined "suitability" in Section 2) is answered just by saying that Φ needs to be high enough, although I do not specifically propose this (and have no argument for it).

Note that the importation of Φ into MDyn would raise the question of what sort of causation is involved in the imported Φ. An immediate possibility would be just to let it be defined as it is at present in IIT, without regard to what MDyn means by "causation" and in particular without involving meta-dynamism (meta-causation). However, a more interesting and coherent (indeed, "integrated") proposal would be, on the one hand, (i) to replace the patterns of causation that IIT relies on by patterns of causation in MDyn's sense, i.e. dynamism at the fundamental physical level, and, on the other hand, (ii) to enrich Φ by having it consider all forms of that dynamism, meta-dynamism included.

Notice that in such an importation of Φ into MDyn we would *not* thereby be importing the panpsychic element of IIT. This element is that, unless one applies a threshold to Φ below which there is no consciousness, it follows that virtually every system has at least some degree and type of consciousness. In the importation, this would no longer be the case even if a threshold were not applied, because the requirements of MDyn might block panpsychism anyway.

Conversely, ideas concerning meta-causation from MDyn might possibly be exported into IIT, provided that the concept could be defined in terms of the type of activity profiles used by IIT in its explication of causation. Thus, the idea is have a "meta" version of causation as conceived by IIT, not to use MDyn's physical meta-dynamism. Then, the IIT meta-causation could, for instance, contribute to the calculation of Φ. If one does not believe that MDyn is on the right lines, one could refrain from imposing meta-causation as a necessary condition for consciousness to exist. But involving meta-causation somehow in IIT could help IIT proponents to defuse some criticisms (see, e.g., [41]), including that IIT is over-liberal in its attributions of consciousness.

5.9.2 Cross-Fertilization with Objective Quantum Collapse Guided by Integrated Information

Kremnizer and Ranchin [33] propose a collapse model (or collapse theory), i.e. a mathematical characterization of a dynamic process whereby the smoothly developing quantum wave abruptly (if still smoothly) collapses. The approach (and other collapse models) integrate collapse seamlessly into overall quantum dynamics rather than leaving it as a separate, unexplained and rather mysterious matter.

Kremnizer and Ranchin's specific proposal involves adding a certain non-linear term to the Schrödinger equation, so that the modified form now encompassed not only the smooth development of the overall quantum wave but also the occasions of collapse. The modified form of Schrödinger equation is schematically as follows [their equation 14):

$$\frac{d}{dt}\rho(t) = -\frac{i}{\hbar}[H, \rho(t)] - \mathcal{I}(\rho(t))$$

where $\rho(t)$ is a density matrix describing the system at hand, the first term on the right is from a standard form of the Schrödinger equation, and $\mathcal{I}$ is a new non-linear operator encapsulating the collapse process (how it arises and what its effects are). The equation follows a general form that Kremnizer and Ranchin report as having been shown to be capable of satisfying certain requirements on collapse models. The main novel contribution is the particular nature of the operator $\mathcal{I}$. Kremnizer and Ranchin make this operator dependent on a measure of integrated information related to IIT's Φ but expressed entirely in quantum-physical terms. The measure, while still labelled Φ, is not intended to be a quantum-theoretic explication of IIT's own measure, but is instead a new measure that is analogous to IIT's Φ. Kremnizer and Ranchin call this information the Quantum Integrated Information (QII) of the system.

Being objectively defined at the fundamental physical level, QII and its use in collapse is more directly relatable to MDyn than IIT's concepts themselves are. We therefore give some further detail of the model to show some opportunities for interaction with the ideas in MDyn. This is despite the fact that the model is not intended as a definite explanation of consciousness, but rather an account of how consciousness can have a role, indirectly, in collapse. On the assumption that consciousness is reflected in QII, consciousness can have a role in collapse to the extent that collapse is controlled by QII. But the model has independent interest as a way of physically explicating collapse, perhaps independently of consciousness.

Kremnizer and Ranchin provide a definition for the QII, notated as $\Phi(\rho)$, of a system in an N-dimensional Hilbert space $\mathcal{H}$ and described by density matrix ρ. Then they define $\mathcal{I}$ as follows in their equation (15):

$$\sum_{n,m=1}^{N^2-1} h_{n,m}(\Phi(\rho(t)))\mathcal{L}_{nm}(\rho(t))$$

Here $\mathcal{L}_{nm}(\rho(t))$ abbreviates an expression that involves $\rho(t)$ and a set of operators on the Hilbert space, determining the basis in which the state collapses. (Kremnizer and Ranchin argue that this should be the position basis in order for the behaviour of macroscopic objects to be explained.) The detailed form and function of the expression need not concern us here. What is of interest is the other, QII-dependent, factor, where the Hermitian matrix elements $h_{n,m}$ are continuous functions that deliver zero when their argument is zero.

Clearly, much as with our comments in Section 4.2 about possible meta-dynamic modifications to the modified quantum theory as proposed by Bohm and Hiley [7], we can equally propose here that, as well as the $\mathcal{I}$ term above, there could be an additional term in the form of a meta-dynamic, gating expression $\mathcal{G}(...)$. Alternatively or as well, a factor in the form of a gating expression could multiply the $\mathcal{I}$ term. One could also consider modifying the h_n, m within the definition of $\mathcal{I}$ by means of gating expressions.

It is less clear how the definition of QII be modified. But, if we are considering a version of quantum theory where dynamism is as real as other aspects of state, then it should presumably be represented in a modified $\rho(t)$ and in a modified Hilbert space $\mathcal{H}$. I leave this as a matter for future research, but any such development would make QII sensitive to dynamism (both base-level and meta-level).

Although MDyn does not itself relate consciousness to collapse (whether to say that consciousness somehow affects collapse, or collapse is constitutively involved in consciousness), there are motivations for MDyn-inspired meta-dynamic modifications to Kremnizer and Ranchin's model. For instance, if MDyn is right that meta-dynamism is necessarily involved in some way in consciousness, and QII measures consciousness, then QII should be suitably sensitive to meta-dynamism.

Conversely, the model, if plausibly correct about how collapse happens as a function of QII, is a candidate for the detailed mathematical framework to use in further development of MDyn, given the comment above that MDyn is most at home with versions/interpretations of quantum theory that collapse collapse into the general dynamics. If, also, QII does measure the strength or richness of consciousness, and/or high QII is a necessary condition for consciousness, it could be given a role in laws separate from the role that the model gives it, for instance by having it as an argument in a new type of gating expression in MDyn.

5.9.3 Orch OR's Objective Reduction

Orch OR [26] proposes that objective quantum-state reduction (collapse), which can "result in moments of conscious awareness and/or choice," takes place in average time τ inversely proportional to the gravitational self-energy E_G of the superimposed states. (This self-energy arises because of different mass-ditributions of the different states in the superposition, resulting in different spacetime curvatures for them.) One immediate suggestion would be to make the avarage collapse time dependent on meta-dynamism as well as gravitational self-energy. But, also, the theory leaves the specific time for collapse arising on a specific occasion to be random. Therefore, there is scope for adding meta-dynamism-dependent refinements to the theory, affecting the detailed timings in a non-random way.

Other possible interactions with MDyn are discernible. MDyn does not propose that consciousness is constitued of, or a result of, collapse events, orchestrated or otherwise. However, suppose that both MDyn and Orch-OR are roughly on the right lines: that some form of meta-dynamism is needed in conscious processes, and conscious processes consist of one or more collapse events. Then, presumably, we must have meta-dynamism playing a role in how at least some collapses happen or proceed. Potentially, this role could be partly or wholly within an individual collapse event, considered as a short-lived continuous process. The process could be meta-dynamically sensitive to its own dynamism. Or, a collapse could even be meta-dynamically sensitive to the dynamism in preceding non-collapse state evolution and/or to prior collapse events.

Importantly, not all collapse events would need to involve (significant) meta-dynamism, and hence it would not need to be the that all have a connection to consciousness. Thus, meta-dynamism could provide an important element of selectivity to Orch OR. The modification would lessen or eliminate the need to talk about all collapse events being, to a small degree at least, occasions of consciousness (events of "proto-consciousness" [26]).

Hameroff [25] proposes that one factor that could usefully regiment the collapses that are involved in consciousness is an "envelope" of a particular sort of neural activity in the brain. Further control, and control of a qualitatively different sort, via meta-dynamism could be a useful adjunct to the envelope idea.

Finally, if Orch OR provides a plausible theory of how collapse works, then, possibly after the subtraction of the claimed relationship between collapse and consciousness, it is another candidate for use as a detailed framework for development of MDyn. Of particular interest here is that this importation would add a form of gravitational dynamism that may otherwise be neglectable. And if collapse is indeed an important component in some types of consciousness, then a conscious process may need to be sensitive to the dynamism involved in the collapsing within that process, and therefore possibly to the dynamism involved in the gravitational aspect of collapse.

6 Conclusion

Various points where further research is necessary or desirable have been noted in passing throughout the article, and I do not summarize them here. Indeed the article sets out a rich framework for further research. Clearly, there is a need for major developments in the mathematical accounts

of dynamism elements, gating expressions, temporal non-locality, the fixed-point equation, auto-sensitivity, and continuity. A big future project is to determine specific changes to specific law-equations. And, of course, as the philosophical side develops this may require major changes on the physical-theory side.

As emphasized in Section 2.3, there is no claim that MDyn rests on definite, unimpeachable arguments that consciousness must involve meta-dynamism. The arguments are merely suggestive, albeit strongly so, I would claim. Even less so is there a definite argument that suitable auto-sustaining, auto-individuating, auto-sensitive meta-dynamism is sufficient for, indeed constitutive of, core consciousness. However, once one has come to these ideas, I would claim that it is clear that they provide an interesting way in which to make sense of the following notion: the notion that phenomenal experiencing is auto-individuating auto-sensing whose core characteristic is directly to sense and individuate that very auto-sensing. It is its own auto-sensing.

Funding: This research received no external funding.

Acknowledgments: I am grateful for the stimulus for pursuing the mathematical side of the work in this article that was offered by being allowed to present a talk in the Models of Consciousness conference at the University of Oxford, UK, 9–12 September 2019.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDyn Meta-Dynamic Theory of Consciousness

Appendix A

A.1 Pre-Reflective Auto-Sensitivity and Direct Acquaintance

The claim that all consciousness constitutively involves pre-reflective self-consciousness has been controversial. Many researchers object that, typically, consciousness, for example visual consciousness during perception, is *transparent*: I am conscious of a red rose, not typically of my consciousness of it. Although I follow Levine [36] in suspecting that such arguments are misconceived (roughly, they do not take into account the phenomenality of the rose-as-seen as itself part of one's consciousness), nevertheless, for safety, I have weakened the assumption to PRAIS, requiring auto-sensitivity without requiring it to be conscious.

One could perhaps describe the assumed auto-sensitivity as a form *direct acquaintance* [27,47] However, this notion is normally at least tacitly assumed to be about conscious acquaintance, so we would need a not-necessarily-conscious version of it. With such a version, the claim would be that a conscious process is necessarily, at each time in its time-span, directly acquainted with its own ongoing unity as a process that is differentiated from the rest of the world. Note that this could then imply that at each moment the process is acquainted with its own past acquaintance instances. (Acquaintance with acquaintance is mentioned in [27].)

A.2 More on Causation and Meta-Causation

In the Introduction to this paper I warn that my usage of the term "causation" refers to something like the productivity or "[causal] oomph" that others have discussed, and not what is predominantly called causation in the philosophy of causation (see, e.g., [16,35]). "Causation" is usually discussed as

being a relation between causal relata that concern separate moments of time and spatially limited objects, events, etc. at those moments, rather than a pushing-forward that exists everywhere at every moment. Those relata are variously held to be a wide range of things, depending on the theory—events, facts, property instances, etc. or types of such things. Also, the theory is typically (though not exclusively) pitched at a level much higher than fundamental physics, for instance everyday human activities, or interactions between simple objects that feature in everyday life (billiard balls colliding, etc.). The notion of causation as being between spatially and/or temporally separated events is also common in discussions about relativity theory, entanglement, and so forth. While researchers often work hard to achieve an objective characterization of causation, in terms such as conditional probabilities, or counterfactual worlds that are supposed somehow objectively to exist, this has certainly been difficult, especially as so often the relata rest on how humans view the world, e.g. how we commonsensically divide the world up into types, or how we take an event to be a particular instance of a particular type. My aim is not to criticize such work but rather just to distance MDyn terminology from how causation is viewed in such studies, and to emphasize that I assume that dynamism is an entirely objective, fundamental physical constituent of the world.[7]

Meta-causation has occasionally been discussed in the philosophical literature. Ehring [16] briefly covers it in an article surveying the philosophy of causation, though he calls it iterated causation (a sub-optimal term as it may sound like chained causation, which is simply where A causes B and B causes C). Koons [31] discusses it under the heading of higher-order causation. However, as Kovacs [32] notes, meta-causation under any name and in anything like the sense I use the term has been remarkably little discussed overall.

There is particular point of possible contact between my (meta-)dynamism and current philosophy. The notion of dynamism has a lot in common with contemporary "power" theory [23,28], where the notion of a power is, or similar to, the notion of a disposition to do something. Dynamism instances may be seen as (partly) constituting what are often called exercisings of powers. Meta-dynamism would become exercisings directly interacting, via their own powers, with other world aspects, including power exercisings. However, to a reader versed in this area, I need to mention that the existing notion of "meta-causal powers" [17], despite its name, does *not* provide the envisaged meta-dynamism/meta-causation acting on power exercisings. Instead, it merely provides operations on other powers as such: it is to do with a power potentially creating, destroying or modifying some *power*, not acting directly upon a power-*exercising*. So they are not a matter of what I would call meta-causation. Klinge [30] discusses,in the context of power theory, a special form downwards mental-to-mental causation that does qualify as meta-causation.

In sum, despite the above connections to the literature, little has been done on meta-causation at any level, and what has been done does not appear to apply to meta-causation in my sense of meta-dynamism. Least of all has meta-causation, whether at a basic physical level or otherwise, been used as a foundation for consciousness (though see some links in [2] to possibly closely related notions). There has been discussion (e.g., in [5,22]) of whether we can be directly, perceptually conscious of causings; and if one holds that perception is partly a matter of things in the world causing happenings in the brain, then one appears to be proposing meta-causation. But this line of thought is at most about how there might be meta-causal influences on our conscious states, not about consciousness itself being made of (suitable) meta-causation, which is what MDyn proposes.

---

[7] Because of this desired distancing, in early work [2] on the current proposal I attempted to avoid the term "causation," instead inventing my own term, "running" or "runningness," to evoke the notion of the universe running systematically like a machine. However, I now replace this by "dynamism." The use of the terminoogy of running is also related to the fact that thinking on MDyn started with wondering about the possibility of conscious AI systems, though always with the intention also of illuminating human consciousness, or consciousness in other living creatures.

A.3 Problems Concerning Representation

As mentioned in Section 2.3, there is, I argue, no theory of representation that is both completely objective and pre-reflective. For instance, in the comprehensive survey of possibilities that Shea [51] goes through in devising a theory of representation, it is possible to discern in every one a point where it either (i) requires something that would be a matter of complex informational structures, or use of concepts, and would therefore go outside pre-reflectivity, or (ii) involves non-objective considerations about what represents what.

The non-objectivity in (ii) arises because, for instance, if the representational relation depends on some sort of structural correspondence between representation and representatee (even one that is matter of logical propositions rather than, say, spatial isomorphism) then there can be no objective, physics-based matter of fact of what the representatee is. This is especially so if we allow for some degree of approximation in the correspondence. If I have a representation in my brain of some some state in my brain, but that state happens to be structurally similar enough to a state in your brain, then I would be as much representing your state as mine: but that cannot work for a theory of *my* consciousness. (See [29] for a related argument.) If one tries to bring in the question of which particular pieces of matter are involved, demanding a particular connection between the piece the representatee is in and the piece the representation is in, then we get into issues of substitution of matter already raised in discussing substrate replacements in Section 2.1, and questions of what are the criteria adequate connectivity between the pieces.

This question of connection also relates to theories of representation that rely in part on the representatee having caused the representation to arise, in a particular, defined way. There are many issues discussed in this area (e.g., see Shea *ibid*) about what one should take the cause in the causation relationship to be. For instance, a dog may somehow cause my representation of the dog, but I could equally say that the light rays etc. between me and the dog, not the dog itself, cause my representation. Similarly, the causation link to my brain typically involves both less than the dog (I'm not receiving much in the way of light rays from its side facing away from me, let alone from its intestines, for example) and more than the dog (I would not be receiving any light rays unless there were a source of illumination). But then there is no completely objective fact of the matter about what is represented. Going back to the states within a conscious process, there would be no objective matter of fact that a current state is indeed representing the particular prior states or state relationships that are allegedly represented.

There are also theories of representation that rely on the representation having historically played a survival-supporting role for that system. But here we have similar issues as regards the objectivity of the causation involved, but also lack of objectivity about how useful the alleged representation has been, how its usefulness is a separate matter from other aspects of the organism, what survival amounts to, etc.

A.4 Conscious Processes as Causally Bound

I claim that consciousness is a property that only a "genuine" process can have. Intuitively, a genuine process is one that is held together by dynamism *within* itself, that gives the process the state trajectory that it has. The contrast here is with *pseudo-processes* (see [15] for review). A common example of a pseudo-process is the shadow of someone walking. States of the shadow do not cause further states of the shadow, under reasonable accounts of causation. Rather, there is a genuine process of someone walking, and causation emanating from each state of that process and leading to changes of illumination that we regard as a moving shadow. The distinction between genuine and pseudo processes is difficult to make precise, partly because we must allow (in general) for even a *genuine* process's unfolding to be affected by "input" consisting of physical interactions with the world outside itself. Fortunately, the present article does not require a precise, objective criterion to exist (let alone

discovered), as the notion is only a way of heuristically guiding argumentation towards the conclusion that meta-causation is a good explanation for the required auto-sensitivity of processes, *whatever* the actual causation it is that binds the process states together.

While it may already be quite intuitive that pseudo-processes could not be conscious, it is useful to have arguments for this. More strongly (or at least more pointedly) I claim that a theory about which processes are conscious cannot be based entirely on considerations about what the mere trajectory of (ordinary) states in a process is like. If it were, it would be possible in principle to "gerrymander" a pseudo-process to have the appropriate trajectory and one would be forced to make very implausible decisions about where consciousness lay in the world. For example, one could gerrymander an inert brain, which would not otherwise be conscious, to have a close approximation to the sequence of states that a brain would go through naturally, by its internal causation, during a conscious experience. The gerrymandering (faking) would be by externally forcing each successive state very rapidly on the overall neural(/chemical/...) network, at some very rapid rate of time-stepping, no matter what the network is itself trying to do (if anything). Then the mere presence of that sequence of states would constitute a conscious process if all that mattered was that sequence, not how the states lead to each other, if the approximation were close enough. However, this can be argued to be implausible. Versions of some arguments related to this are given in [2], but the point relates to many thought experiments (see, e.g., [6,29]) about the nature of consciousness.

One side of such argumentation is that, if the successive states of a pseudo-process alleged to be conscious do not lead naturally to each other, why it matters at all that they occur in the *temporal order* that they do. Or for that matter, they they occur in an order at all, as opposed to being simultaneously realized on numerous isomorphic brains. After all, many thought experiments about consciousness rest on the idea that if one swapped out some neurons and replaced them by other physically isomorphic ones, it would not make any difference to the conscious process. And of course we need to allow a conscious process to move around in space. So, combining these observations, the fact that a state appears on some set of neurons somewhere and the next state appears on another, isomorphic set of neurons somewhere else (even in a different brain) is totally irrelevant to the existence of the consciousness, if such existance is not conditioned on causal binding.

A variant of this line of thought has the original successive gerrymandered states being put on different brains but keeping their time order. So brain N has state N at some time. But if the resulting process is conscious, as it should be if causal links and which-particular-bit-of-matter do not matter as regards the existence of consciousness, then we should note the following. We would be enabled to say that, for instance, when there are N conscious people all pursuing identical physical trajectories inside their brains, then, as well as these N conscious episodes, there also hugely many other ones formed by choosing successive states of a process from *different* brains. Part of the point here is that there is no objective criterion for deciding what sequences of states one should consider.

In [2] I also pursue the related argument that the kth brain could just permanently, statically hold the kth state, and an observer just picks a brain at each time step by some external physical means, for instance by shining a light on it. This way, we have a temporal sequence of the required states (albeit that each required state is embedded in a bigger one that involves the light at a particular position). So why should not that the resulting overall process be conscious? But do we really want to say that consciousness arises just by the process of moving the light? If it doesn't arise, why not?

Of course, the involvement of brains in these comments is only incidental. Similar comments could be made about any type of system that one was considering to be bearers of consciousness, for instance computers running programs.

In the above we have not needed to rely on exact gerrymandering of trajectories, exact isomorphism etc. in the arguments, because we are allowing, as I believe we should, a degree of "transverse continuity" in the sense of Section 4.6. And this addresses a fundamental issue as regards gerrymandering. If one does not demand transverse continuity, and if one assumes that spacetime is continuous, one my go on to claim that no amount of departure from real conscious process trajectories, no matter

how tiny, can be tolerated in gerrymandering. But with the continuity assumption, one can imagine a form of gerrymandering where a discretized version of the continuous state trajectory of a process is what the gerrymandering produces, where between time instants t_n the state is kept constant (say). Then with sufficiently fine discretization, one can get as close as desired to the original continuous trajectory. This makes gerrymandering thought experiments more like something one could imagine could happen in practice (at some time in the future), and makes it acceptable to even be thinking about a discrete sequence of states (as for instance imagining a kth state on a kth system).

A.5 An Argument for Meta-Causation

With arguments such as those alluded to in the previous section, one might be convinced *that* a characterization of conscious processes must allude to their internal causation, not just to their state trajectories. But then one can ask: *in what way* does the described causation matter? What jobs is it doing in the constitution of consciousness, over and above any job being done by the states that the causation binds together?

My intuitive answer here is that: *the causation matters to the conscious process itself.* This is at the nub of the assumption of pre-reflective auto-sensitivity and the argument from it to meta-causal auto-sensitivity. Crucial also is that this mattering-to-itself needs to *entirely objectively* exist: it cannot have any whiff whatever of alternative decisions a person might make about what the state of the world is. This is because (I assume) consciousness is an entirely objective, physical property of physical processes. To put it another way, mattering-to-itself is the only way of getting some physically objective sort of mattering, because otherwise one would have to be proposing something else to which it mattered: that other thing had better not be a conscious system, but on the other hand it is difficult to see why any sort of mattering to a non-conscious thing should contribute to making our original process conscious.

A particular argument that that intuitive line of thought led to is as follows. I assume that it has been established that the internal causation within processes matters to the question of whether they are conscious, that PRAIS is true, and furthermore that representation within a given process state of past states and/or their causal binding cannot satisfy PRAIS. Let us suppose we are alternatively investigating whether some sort of direct causation from the history of the process to an effect on a current state will provide PRAIS.

Then: let us suppose, for a *reductio ad absurdum*, that that causation is only from *states* within the history, not from the causation within it. We can even allow here that a whole (sub-)trajectory could have an effect different from the combined effect of the individual states in that trajectory. I am not proposing that this collective added-value could in fact happen, but just allowing it for the sake of argument, in case it were to be claimed that auto-sensitivity with respect to state-collections rather than to causation is adequate for consciousness. The *reductio* goes as follows.

Consider the state P_t of a conscious process P at time t.[8] Consider its prior state trajectory, $P_{<t}$. This plus the state P_t has all the needed pre-reflective auto-individuating auto-sensitivity, by assumption, since P is conscious. I assume in the following that this means that $P_{<t}$ is a conscious process that is an initial portion of P. Suppose that auto-sensitivity is achieved, as stated above, by direct causation from past states, so that in particular the auto-sensitivity at t is achieved by direct causation from states in $P_{<t}$ to the current state P_t. Now imagine a trajectory of states elsewhere in space, but converging on P's spatial location at t, that is a (near-)isomorphic to $P_{<t}$, but that does not contain the auto-sensitivity required for consciousness. For convenience we can call this trajectory $Q_{<t}$. We also imagine that, whatever direct causation from $P_{<t}$ states exists into P_t, (near-)isomorophic

[8] I am assuming here that it is an "ordinary" sort of state, in the sense used in the main text. The current argue is of course on a path toward proposing meta-dynamism and reified dynamism, and so cannot rely on those concepts.

causation exists from the corresponding states in $Q_{<t}$ into P_t. But, even if the effects of $P_{<t}$ states and $Q_{<t}$ states on P_t are slightly different, there is no objective criterion for saying that P_t individuates $P_{<t}$ as opposed to $Q_{<t}$ as being part of the experience that P_t is within. Both of theose trajectories have exactly the same type of relationship to P_t, because P_t is entirely unaffected by what causation is present in the trajectories. Hence, any condition we might propose that would explain P's auto-individuation at time t as having (a recent part of) $P_{<t}$ as its history would also say that P individuated itself as having (part of) the history $Q_{<t}$ as part of itself. Thus, there is no adequate auto-individuation. Therefore, PRAIS is not satisfied in P. This is a contradiction.

We allow mere near-isomorphism in the argument so that it does not have to rely on the idea that $P_{<t}$ can be exactly copied or that we need to copy the state at every instant within it.

But in fact allowing divergences of $Q_{<t}$ from $P_{<t}$ is also useful to the argument in another way. A conceivable line of objection to the argument is that if the two candidate histories are identical, it does not matter if P makes the wrong choice, so to speak, or wrongly includes both in itself, because in effect it is still correctly identifying its history up to state-wise isomorphism.

In fact, the argue doesn't need $Q_{<t}$ to be similar $P_{<t}$ at all, except that $Q_{<t}$ should be a possible history of *some* conscious process that has state P_t at time t, and that the effect it has (via the direct causation being assumed) on P_t is the same as the effect it would have had if it had actually been P's history before t. The only reason for making $Q_{<t}$ very similar to $P_{<t}$ was to have a particular way of creating such a possibility.

Now compare the scenario imagined in the argument to one where there is a $Q_{<t}$ as as above except that it, plus P_t, *is* conscious and does contain the necessary auto-sensitivity for consciousness. Then in fact we have a situation where the overall conscious process is not, after all, just P, but is P with $Q_{<t}$ added as an extra strand of history. This is perfectly accapatble from the point of view of our auto-sensitivity requirements for consciousness: our treatment does not require the auto-sensitivity at any time to be sensitivity to the dynamism across all spatial positions at earlier times, so it is possible for the $P < t$ strand to have auto-sensitivity only with respect to itself, and simply for the $Q_{<t}$ strand. The scenario can also consistently be described as one where there are two separate consciousnesses before t that merge at t to form one.

Bearing in mind that it is possible that the auto-sensitivity in consciousness may just stretch over a sliding window rather than encompass the whole past history of the process (Section 2.4), it may be that at times sufficiently after t, P is not "aware" of its forked history in any sense at all. But if it does have complete historic auto-sensitivity, then it is still not assumed that P senses the forking as such (and least of all that P at or after t, even if capable of states such as belief, consciously realizes that it has resulted from merging). It may just sense the two strands a combined unity.

Returning to the above conceivable objection, an important relevant point to note is that, despite the assumptions of mobility and portability across substrates in Section 2.1, different but absolutely isomorphic processes at different locations form numerically distinct consciousnesses. For instance, imagine two brains that happen to be supporting exactly isomoprhic conscious processes. I claim that we still have two consciousnesses, not one, even if it would be reasonable to claim that they each felt the same to themselves. So even if $Q_{<t}$ were identical to $P_{<t}$ in the above *reductio*, it is wrong for P to include $Q_{<t}$ in its auto-individuation, because even it were conscious (which it is not) it would be a different consciousness from $P_{<t}$.

There are other possible lines of objection to the argument. For instance, it might be claimed that even the slightest divergence of $Q_{<t}$ from $P_{<t}$ would render it implausible as a conscious history in the first place, so there might be a basis for saying that P objectively has a way of including $P_{<t}$ in its auto-individuation but not $Q_{<t}$. A more general objection that encompasses that one is that it might be claimed that *any* trajectory lacking the auto-sensitivity needed for consciousness has systematic, noticeable differences in terms of its states from *any* trajectory that does have the auto-sensitivity. Against this objection, we have the following. There is no reason to think that, given a conscious process, a gerrymandered near-copy of it, as close as we like to the original in terms of of state trajectory,

but not containing the auto-sensitivity, cannot exist, but on the other hand there are positive reasons for thinking it could exist. I will just give a suggestive illustration on this point.

Suppose a conscious machine were constructed, laid out on a 2D plane, and where the relevant state at each spatial point for the conscious process was how brightly it is shining. The operation of the machine cause the brightness pattern to vary in time, where this causation contains all the auto-sensitivity required for consciousness. Now consider a film of this machine, and a projection of the film on a screen somehwere else. It has exactly (or as exactly as one might wish) the same sequence of brightness patterns as the original machine, but not only lacks the auto-sensitivity needed for consciousness but is denuded entirely of the internal causation that we presume is needed for consciousness. The projected pattern is a pseudo-process highly akin to the walking person's shadow in Section A.4. This illustration goes beyond the exact mathematical treatment in Section 4 of the main text, by not including all physical state at a given location in the original machine, but we have already said it is desirable to generalize the treatment to cope with such partialness.

Another, perhaps simpler, point is that conscious processes will generally receive "input" from their environments. Consider two conscious processes that are on identical brains or machines and start off identically, but receive slightly different trajectories of input. So, the courses the processes take over time are very similar. Someone who claimed that inlcuding a type of auto-sensitivity necessarily makes a noticeable difference to trajectories would have to claim that this difference was unlike any difference that could possibly arise from differences merely in the input. This seems like a difficult condition to meet.

References

1. Adlam, E.(2018). Spooky action at a temporal distance. *Entropy, 20*, 41 [online]; doi:10.3390/e20010041
2. Barnden, J.A. (2014). Running into consciousness. *J. Consciousness Studies, 21* (5–6), pp.33–56.
3. Barnden, J.A. (2018). Phenomenal consciousness, meta-causation and developments concerning casual powers and time passage. Poster presented at *22nd Conference for the Association for the Scientific Study of Consciousness,* 26–29 June 2018, Kraków.
4. Barnden, J.A. (2019). Consciousness and meta-causation. Talk at *Joint Session of the Aristotelian Society and the Mind Association*, University of Durham, UK, 19–21 July 2019.
5. 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.
6. Bishop, J.M. (2009) Why Computers Can't Feel Pain. *Minds and Machines,* 19, 507-516, (2009).
7. Bohm, D. & Hiley, B.J. (1993). *The undivided universe: An ontological interpretation of quantum theory.* London and and New York: Routledge.
8. Brüntrup, G. & Jaskolla, L. (Eds) (2016). *Panpsychism: contemporary Perspectives*. Oxford Scholarship Online, October 2016.
9. Chiou, D.-W. (2015). Loop quantum gravity. *Int. J. Mod. Phys. D, 24*(1).
10. Cucu, A. & Pitts, B. (2019). How dualists should (not) respond to the objection from energy conservation. *Mind and Matter, 17*(1), pp.95–121.
11. Carruthers, P. (2011) Higher-order theories of consciousness. In E.N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2011 ed.).
12. Cole, D. (2020). The Chinese Room Argument. In E.N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2020 ed.).
13. Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness.* New York: Harcourt Brace 1999.
14. Demarest, H. (2017). Powerful properties, powerless laws. In J.D. Jacobs (Ed.), *Causal Powers,* Chapter 4. Oxford: Oxford Scholarship Online.
15. Dowe, P. (2009). Causal processes. In *Stanford Encyclopedia of Philosophy,* Spring 2009 edition.
16. 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.

17. 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.
18. Ellis, G. (2016). *How can physics underlie the mind? Top-down causation in the human context.* Dordrecht: Springer.
19. Gallagher, S. & Zahavi, D. (2015). Phenomenological approaches to self-consciousness. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition).
20. Gellman, J. (2017). Mysticism. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition).
21. Ghirardi, G. (2018). Collapse theories. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition).
22. 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.
23. Groff, R. & Greco, J. (Eds) (2013). *Powers and capacities in philosophy: the New Aristotelianism.* New York and London: Routledge.
24. Guillot, M. (2017). *I me mine:* on a confusion concerning the subjective character of experience. *Review of Philosophy and Psychology, 8*, pp.23–53.
25. Hameroff, S. (2010). The "conscious pilot"—dendritic synchrony moves through the brain to mediate consciousness. *J. Biol. Phys., 36*, pp.71–93.
26. Hameroff, S.R. & Penrose, R. (2016). Consciousness in the universe: an updated review of the "Orch OR" theory. In R.R. Poznanski, J.A. Tuszynski & T.E. Feinberg (Eds), *Biophysics of Consciousness: A Foundational Approach*, pp.517–599. Singapore: World Scientific.
27. Hasan, A. & Fumerton, R. (2017). Knowledge by acquaintance and knowledge by description. In E.N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2017 Ed.).
28. Jacobs, J.D. (Ed.) (2017). *Causal powers.* Oxford: Oxford Scholarship Online.
29. Kirk, R. (2005). *Zombies and consciousness.* Oxford: Clarendon Press (Oxford University Press).
30. Klinge, F. (2019). The role of mental powers in panpsychism. *Topoi*, published online 19 Jan 2019.
31. Koons, R.C. (1998). Teleology as higher-order causation: A situation-theoretic account. *Minds and Machines, 8*: 559-585.
32. Kovacs, D.M. (2019). The question of meta-causation. Talk at *FraMEPhys/Gothenburg Conference on Metaphysical Explanation in Science,* co-organized by Philosophy department, University of Birmingham. Birmingham, 10–11 Jan 2019.
33. Kremnizer, K. & Ranchin, A. (2015). Integrated information-induced quantum collapse. *Foundations of Physics, 45*, pp.889–899.
34. Kriegel, U. (2009). *Subjective consciousness: A self-representational theory.* Oxford University Press.
35. Kutach, Douglas (2014). *Causation.* Cambrige, UK: Polity Press.
36. Levine, J. (2018). *Quality and Content: Essays on Consciousness, Representation, and Modality.* Oxford Scholarship Online: April 2018.
37. Maudlin, Tim (2007). *The metaphysics within physics.* Oxford: Oxford University Press.
38. McQueen, K. (2017). Does consciousness cause quantum collapse? In issue on Radical Theories of Consciousness, *Philosophy Now, 121*, August/September 2017.
39. Metzinger, T. (2018a). Minimal phenomenal experience – a new theory about pure consciousness "as such." Talk at *22nd Annual Meeting of the Association for the Scientific Study of Consciousness* (ASSC 22) Jagiellonian University, Kraków, Poland, 26–29 June 2018.
40. Metzinger, T. (2019). From MPS to MPE: Meditation, tonic alertness, and minimal phenomenal experience. Tutorial at *23d Annual Meeting of the Association for the Scientific Study of Consciousness* (ASSC 23), Western University, London, Ontario, 25–28 June 2019.
41. Mindt, G. (Ed.) (2019). Special issue: Reflections on Integrated Information Theory. *J. Consciousness Studies, 26*(1–2).
42. 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.
43. Pylkkänen, P. (2018). Quantum theories of consciousness. In R.J. Gennaro (Ed.), *The Routledge Handbook of Consciousness,* pp.216–231. New York and London: Routledge.

44. Real analytic function. *Enclyclopedia of Mathematics.* http://www.encyclopediaofmath.org/index.php?title=Real$_$analytic$_$function&oldid=31091
45. Rosenberg, G. (2016) Land ho? We are close to a synoptic understanding of consciousness. In G. Brüntrup & L. Jaskolla (Eds), *Panpsychism: Contemporary Perspectives*. Oxford Scholarship Online, October 2016.
46. Rosenthal, D.M. (1993). State consciousness and transitive consciousness. *Consciousness and Cognition, 2* (4), pp.355–363.
47. Russell, B. (1910) Knowledge be acquaintance and knowledge by description. *Proceedings of the Aristotelian Society, 11,* pp.108–128.
48. Schaffer, J. (2016). The metaphysics of causation. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition).
49. Searle, J. (1980) Minds, brains and programs. *Behavioral and Brain Sciences*, **3**, pp. 417–424.
50. Sebastián, M.A. (2012). Experiential awareness: Do you prefer *it* to *me*? *Philosophical Topics, 40*(2), pp.155-177.
51. Shea, N. (2018). *Representation in cognitive science.* Oxford: Oxford University Press.
52. Wallden, P. (2010). Causal Sets: Quantum gravity from a fundamentally discrete spacetime. *J. Physics: Conference Series, 222*: 012053.
53. Weinstein, S & Rickles, D. (2019). Quantum gravity. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy,* Summer 2019 ed.
54. Wiese, W. (2019). Explaining the enduring intuitn of substantiality: The phenomenal self as an abstract 'salience object'. *J. Consciousness Studies, 26*(3–4), pp.64–87.
55. Williford, K. (2015). Representationalisms, subjective character, and self-acquaintance. In T. Metzinger & J.M. Windt (Eds), *Open MIND*: 39(T). Frankfurt am Mein: MIND Group.
56. Windt, J. (2015). *Dreaming: A conceptual framework for philosophy of mind and empirical research.* Cambridge, MA: MIT Press.
57. Windt, Jennifer (2019). Consciousness in dreams and dreamless sleep. Keynote Lecture at *23d Annual Meeting of the Association for the Scientific Study of Consciousness* (ASSC 23), Western University, London, Ontario, 25–28 June 2019.
58. Windt, J., Nilesen, T. & Thompson, E. (2016). Does consciousness disappear in dreamless sleep? *Trends in Cognitive Sciences, 20*, pp.871–882.
59. Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective.* Cambridge, Mass. and London, UK: MIT Press (a Bradford Book).