Review of Philosophy and Psychology
https://doi.org/10.1007/s13164-021-00603-z

Pre-Reflective Self-Consciousness: A Meta-Causal Approach

John A. Barnden[1,2]

Accepted: 22 November 2021

Abstract
I present considerations surrounding pre-reflective self-consciousness (PRSC), arising in work I am conducting on a new physicalist, process-based account of [phenomenal] consciousness. The account is called *the meta-causal account* (MCA) because it identifies consciousness with a certain type of arrangement of meta-causation. Meta-causation is causation where a cause or effect is itself an instance of causation (e.g., A's-causing-B causes C). The proposed type of arrangement involves a sort of time-spanning, internal reflexivity of the overall meta-causation. I argue that, as a result of the account, any conscious process has PRSC. Hence, PRSC does not need to be taken as a stipulation or argued for on purely phenomenological grounds or as a necessary support for reflective consciousness. I also show how it is natural to the account that PRSC is not an additional, peripheral, sort of consciousness, but is intrinsic to all consciousness, thereby fitting claims about self-intimation and co-constitution by various authors, and also being amenable to an adverbial account. As part of this, consciousness of an external object is just the form that current self-consciousness takes, the meta-causal constitution of it being inextricably modulated by the causal relationship with the object. The article also discusses how MCA helps explore issues of for-me-ness, transparency of perceptual consciousness, and possible immediacy and non-relationality of self-consciousness.

1 Introduction

This article claims, in brief, that *pre-reflective consciousness is physically constituted of meta-causation*. This view provides a new, fruitful, physicalist basis for exploring matters like the intrinsicality of self-consciousness to consciousness, a self-intimation quality of consciousness, for-itself-ness (and for-me-ness) of consciousness, and apparent transparency of perceptual consciousness. In particular,

✉ John A. Barnden
jabarnden@btinternet.com

[1] School of Computer Science, University of Birmingham, Birmingham, B15 2TT, UK

[2] FraMEPhys Research Group, Department of Philosophy, University of Birmingham, Birmingham, B15 2TT, UK

Published online: 22 January 2022

the view provides an unusual argument for the existence of pre-reflective self-consciousness (PRSC), not relying on phenomenological intuitions or the idea that it is a necessary support for reflective self-consciousness.

The concept of meta-causation, in this article's sense,[1] is far from well-known. By way of an initial guiding intuition, an instance of meta-causation is a causation instance where something on the cause side and/or something on the effect side is itself an instance of causation, as in the pattern [A-causes-B]-causes-C, or A-causes-[B-causes-C], or [A-causes-B]-causes-[C-causes-D] (and similarly with additional partial causes or effects included). If someone said *"John's causing Bill to cry caused Mary to get angry"* they would ostensibly be describing a meta-causing (meta-causation instance). Of course, the mere existence of such sentences is not evidence that even causation, let alone meta-causation, exists objectively in the world, rather than being a matter of human construal. However, this article does claim that there is real physical causation and real physical meta-causation in the world. Also, its preferred view is that causation is an aspect of the most fundamental layer of the physical world, rather than naturally sitting at the everyday-world level of people, crying, anger, etc.

Only a small minority of mainstream philosophical research on causation even mentions meta-causation (as affirmed also by Kovacs 2019, 2021), so talking about it at all is fairly radical. Proposing it as physically real is even more radical, and proposing it as the basis for consciousness is yet more so. But meta-causation has received occasional mention, for instance in Ehring (2009), where it is called iterated causation, as it is also in Kovacs (2021), though it must not be confused with mere chained causation (A causes B, B causes C). A type of meta-causation is central in Koons (2009), where it is called higher-order causation. Celestial meta-causation in the form of a god continuously causing the physical causation in the world (mentioned briefly by Kovacs 2021) is one nexus of discussion of meta-causation, one which is, however, not relevant to this article.

What has meta-causation to do with the fundamental nature of consciousness? (By consciousness I will always mean phenomenal consciousness). This article's account, called here MCA (Meta-Causal Account), gets to meta-causation from consciousness by the following route. It first argues (as many have) that consciousness is based somehow on reflexivity—some sort of access the conscious system has to itself. MCA is process-based, and argues that this access consists in a conscious physical process *being sensitive to its own process-ness as such.* MCA takes this auto-sensitivity to be a matter of the process *being sensitive to its own internal causation,* i.e., sensitive to, as a separate entity in its own right, the stream of causation that leads this very process through its state trajectory. MCA then argues, because of a desire for pre-reflectivity and absolute objectivity of the auto-sensitivity, that the auto-sensitivity cannot plausibly be a matter of the process *representing* its own causation. Instead, it proposes that the auto-sensitivity is *meta-causal:* the process's

[1] The term is sometimes used instead for downwards causation, including from mental states to physical states, or for causal influences on the parameters of a causal model. Also the "meta-causal powers" of Ellis (2013) are not a matter of meta-causation in my sense, but instead are causal powers that affect causal powers.

internal causation, as a "new" sort of physical entity in its own right, causes effects within the process.

These considerations have led me to the claim, in MCA, that a specific sort of arrangement of meta-causation is *necessary* for a process to be conscious. However, MCA then adds the conjecture that such an arrangement is also *sufficient* for pre-reflective consciousness, and indeed *constitutive* of it. MCA is therefore a physicalist type-identity theory of consciousness: consciousness as a feature of a physical process is identical to the special sort and arrangement of physical meta-causation lying within the process.[2]

I should stress that *there may possibly be other meta-causation in the world with no relevance at all to consciousness.* MCA, having introduced meta-causation, has no reason to restrict it to lying within consciousness, even though I currently have no specific argument or evidence for positing some other sort of meta-causation. Consciousness involves a *special arrangement of* meta-causation, not just any old meta-causation.

I wish to allow for the possibility that human babies and organisms that are way below the level of human beings have some form of phenomenal consciousness, be it only in very crude forms of pleasurable and painful feeling, for instance. I do not positively claim such organisms are conscious, but I wish to avoid ruling this possibility out of court by virtue of the cognitively advanced nature of mechanisms and processes that a consciousness theory might propose.

MCA, being thoroughly physicalist, takes the property of being conscious to be a purely and thoroughly objective matter, free of aspects based on non-objective human ways of construing the world. This objectivity has played a strong role in the development of MCA. (By "objectivity" I am not directly alluding to the distinction made between "subjects" and "objects" of consciousness. For instance, MCA allows "subjects" in some form to exist objectively, although the issue is hardly touched on in this article.)

The paper is structured as follows. Section 2 discusses why one should think consciousness intrinsically involves some sort of reflexivity, and argues that conscious processes are physically sensitive to their own internal causation. Section 3 further summarizes MCA and the remaining reasoning leading to it. Section 4 argues that the existence of PRSC follows naturally in MCA, discusses how outer-idrected consciousness in MCA is a flavoured form of auto-consciousness, and further explains the nature of reflexivity in MCA. Section 5 presents other consequences and ramifications mentioned above, concerning intrinsicality, self-intimation (focussing on work by Galen Strawson), for-me-ness and the appearance of transparency. Section 6 presents some additional discussion, on MCA's avoidance of panpsychism (here diverging from Strawson) and MCA's possibly having instead a quality of "bathypsychism" (e.g., consciousness in suitable micro-physical systems). Section 7 concludes.

But first, some preliminary notes.

[2] In MCA, consciousness is NOT fundamentally a matter of a physical process's causal role with respect to the world *outside* itself. However, as we will see, incoming causation from the outside world governs the question of what external matters the process can be said to be conscious of. Similarly, outgoing causation governs the question of what external things the process can be said to be consciously making happen, though this issue will not be addressed further in this paper.

I usually replace "self-" by "auto-", for instance replacing "self-consciousness" by "auto-consciousness". I accordingly replace the standard term "pre-reflective self-consciousness" (PRSC) by *"pre-reflective auto-consciousness"* (PRAC). The prefix "self-" is notoriously ambiguous, at times referring to some sort of holder or agent of an experience, possibly a stable entity that has a stream of experiences over a long period of time, and at other times being merely a linguistic prefix tantamount to "itself", as in "self-cleaning oven". But "auto-consciousness" just borrows the latter sense, without, however, precluding the possibility that there is *also* a self somehow associated with the conscious episode.

The "pre-reflectiveness" or "non-reflectiveness" of a mental state is commonly taken to imply that it is cognitively basic and does not use concepts, descriptions, propositional constructions and the like. I adhere to this in this article. In my brief mentions of reflective mental states, I assume that they do typically involve concepts, etc. I do not engage with the important matter of whether reflective states do, and pre-reflective states don't, involve a deliberate turning of mental attention to a target state, and the matter of whether such deliberateness is conscious or not. I don't assume that reflection implies consciousness, so will distinguish between conscious and unconscious reflection as necessary. For convenience, I use "reflective" and "pre/non-reflective" to qualify not just types of auto-consciousness but also types of *outer-directed consciousness*—consciousness of external matters. The distinction thus being made is whether that consciousness involves concepts, descriptions, etc.

What I will call the "full" MCA includes a commitment to the particular physical nature of causation and meta-causation, whereas the present article leaves this nature open, as this nature does not affect the article's concerns, except in comments in sections 6. That physical nature is given a mathematical regimentation in Barnden (2020), where causation is called dynamism and the full MCA is called MDyn (for meta-dynamism). But that treatment does not need to be considered in order to understand the present article.

2 The Reflexivity of Consciousness or of its Underpinnings

2.1 Noticing or Inferring Reflexivity

Why think that consciousness involves any type of reflexivity? Reflexivity here means access by one's mental states to one's mental states, with a possible strong special case being access by a given state to itself. Further, one might be interested in the sub-question of why one should think that consciousness involves auto-*consciousness*, as opposed to some sort of *unconscious* access by the organism to its own mental states. One answer to the sub-question is a common intuition that our consciousness constantly involves some sort of basic, background consciousness of that consciousness (see, e.g., Kriegel 2009). But some philosophers say they don't share this intuition (see, e.g., Gennaro 2018). There is also the (introspective) observation that we often introspect—often reflect consciously, and often in considerable detail, on our own conscious thoughts, sensations, feelings, etc. Such introspection is straightaway a reflective form of reflexivity, of course. But then, more interestingly,

it has been argued that such reflection requires some already-existing, very basic, unmediated, *non*-reflective reflexivity, leading to the supposition that there is non-reflective auto-consciousness. (See, for instance, Frank 2007, 2016, 2019, who emphasizes J.G. Fichte's work [Fichte 1982] as a prime origin of the argument.).

However, in developing MCA, I wished, on the one hand, to be careful not simply to agree with the above phenomenological intuition and summarily shove aside doubters, and I wished, on the other hand, to be guarded about how much can be inferred from the ability consciously to reflect on one's mental states. On the latter point, it seems fine to infer, from the ability to reflect consciously on one's (conscious) mental states, that there is some sort of pre-existing, non-reflective reflexivity (auto-sensitivity or auto-access). But I do not see that it easily follows that that reflexivity should amount to some sort of auto-*consciousness.* This worry arises for me because, as part of views about the mind that form the backdrop for MCA, I assume that there can be unconscious mental states, and that they can be perfectly reflective in the sense of involving concepts, propositions, etc. Given that, I see no reason why there cannot be reflective unconscious thoughts that are about one's other (unconscious or conscious) mental states, or that are even about themselves.[3] Thus, the argument that is used about *conscious* reflection about one's mental states—to infer some sort of non-reflective unmediated access to them as the foundation for that reflection—applies just as much to *unconscious* reflection on them, allowing us to infer non-reflective unmediated access as a foundation for *that* reflection. It would be otiose at this point to take this access to be in any way conscious. But then, it seems such access might also be adequate as the foundation for conscious reflection as well—the access enables the conscious mind to "find" its own mental states, so to speak, though only doing this behind the scenes, and consequently to present the *found* states to reflective consciousness. It is not at all clear that the found states absolutely need to have been subject to some sort of *conscious* apprehension all along (although such apprehension would arguably have been *sufficient* for enabling the reflection).

In sum, while I share the intuition that consciousness involves some sort of reflexivity, and my worries in the previous paragraph are not decisive, they present enough of a problem to explain my not starting from an assumption of auto-*consciousness*. Rather, I claim one does better to rest on reflexivity that is not assumed from the start to be reflexivity of consciousness—though it may then turn out to be such, through further argument or empirical discovery.

[3] In higher-order thought (HOT) theories, a creature's mental state which would otherwise be unconscious is conscious when it is being represented by a higher order state belonging to the creature, but only if certain extra conditions are met, such as non-inferentiality of the inter-state connection. (See, e.g., the Extra Conditions problem in Van Gulick 2006.) The higher-order state can itself be unconscious (and usually would be). So even a HOT theory can allow for creatures unconsciously reflecting on their unconscious mental states, when the extra conditions are unmet.

2.2 The Auto-Mattering of Process-ness

The worries in section 2.1 aside, one aspect of MCA is a positive argument, or at least motivational line of thought, in favour of a form of reflexivity underlying consciousness that does not assume or obviously imply consciousness of consciousness. It is expressed in terms of physical causation, and is as follows.

First, the argument rests on a prior assumption that it is physical processes that are the possible bearers of consciousness. I cannot go into the reasons for this assumption, but it at least roughly reflects process philosophy (Seibt 2013), and accords with one strong trend in consciousness theory, old and new, namely that consciousness is fundamentally a matter of activity rather than static state. To take a few diverse examples, it accords in this respect with the Integrated Information Theory of consciousness (Oizumi, Albantakis & Tononi 2014), the centrality of activity in the work of J.G. Fichte (see Fichte 1982), and the explicitly processual basis of Galen Strawson's work (2017: chs1,3). In particular, Strawson (2017: ch.8) affirms the need for something dynamic in the reflexivity of consciousness, rather than merely static reflexivity as in some linguistic items (e.g., "this very sentence"). Various other theories of consciousness, while not focused explicitly on processes, give an important role to activity, such as the WIV (Wide Intrinsicality View) of Gennaro (2012, 2016). Gennaro (2012: 95, 100; 2018) casts the "meta-psychological" states as higher-order conceptual activity in dynamic integrative interaction with the lower-order states they represent. The HOGS (Higher Order Global State) theory of Van Gulick (2006) is very heavily activity-based, to the extent that its account of representation and self-representation is intrinsically a matter of complex, world-engaged activity. See also Van Gulick (2010–2011) for the "teleo-pragmatic" theory of mind in which HOGS is embedded, the "pragmatic" referring to activity in the world. Van Gulick (2006) says that philosophers sometimes focus solely on the static or synchronic features of consciousness, "but its dynamic diachronic flow is also a central feature of how we experience ourselves and our world." As we will see, MCA integrates parts (interim histories) of a conscious process into the whole process. This integration is very different in detail from HOGS's integration of lower-order states of activity into more global ones, but chimes with the latter's emphasis on integration into a flow of activity.

In addition, and again in line with Strawson's work (Strawson 2017: ch.10), MCA has it that *consciousness is not defined at instants:* consciousness is intrinsically a matter of what happens over some time interval of positive extent, even if very short. It makes no sense to say that a process (or whole person, etc.) is conscious at time instant T, except as a way of saying that the process etc. is conscious over some interval that includes T. Moreover, it is not that there is a chain of successive intervals, one ending where the next one starts, but rather a sliding window. Successive versions of the window overlap, knitting together a united conscious process. (Cf. Schechter 2018 on the diachronic unity of consciousness possibly being a matter of such an overlapping succession.)

I take a process to be a chunk of spatially-extended causally-unfolding activity of the world over some time. The process unfolds largely because of its own internal causal linkage between the states the processes traverse. But this progression

is typically helped along by causation "coming into" the process from the physical surrounds of the process. For convenience I talk both of the former and of the latter types of causation (plus any causation going out of the process into its surrounds) as being "within" the process. I say the causation is "fully" or "entirely" within or internal to the process when I wish to exclude the incoming or outgoing causation. More conveniently, causation within the process can be called the process's *causal spine*. The entirely-within aspect is the *narrow causal spine*.

When I mention a conscious process I mean one that is uninterruptedly conscious: every sub-interval of the process's time-span is fully occupied by a conscious sub-process (unless the interval is too short to sustain consciousness at all, should there be a minimum temporal extent the physical underpinnings of consciousness need to have in the type of system in question). I also oversimplify a little in this article by using "conscious episode" to mean an uninterrupted conscious process, putting aside the possibility that two or more conscious processes with a temporal gap in between could still feel like one uninterrupted period of consciousness.

Given this background, MCA takes the following to be plausible: *consciousness is not constituted merely by the process going through the particular sort of sequence of instantaneous states that it goes through. It is crucial that those states arise in a causally coherent way.*[4] I will mainly have to leave this as a working assumption in the present article, in the hope that it has intuitive appeal, with some readers perhaps even finding it too obvious even to mention. [5]

But now a fundamental question arises. Given that the causation within a process, not just the consequent trajectory of states, matters as regards the process being conscious, *why* should it *matter?* What *difference* does the causation make, over and above supporting the state trajectory? The claim in the previous paragraph says that it *does* matter, but doesn't explain what the causation in itself actually contributes to the constitution of consciousness. We could leave it that its mattering is just a "brute" fact, but it is worth seeing whether we can escape this brutality. It would be insufficient for it only to matter to a human observer, because then consciousness would become non-intrinsic to the process and a non-objective matter of construal. I suggest therefore that the causation in a process matters to its being-conscious *because the causation matters to that process itself* (in a way that goes beyond the sheer fact that the causation is what binds the process states together, as is the case also in non-conscious processes). This mattering is not a question of the process conceptually thinking about its own internal causation. Rather, the process somehow *detects (part of) its own causal spine as an entity in its own right*. Because the

[4] I do not take the division of the world into processes to be an objective matter. But when we are theoretically considering a particular chunk of the unfolding world as a process, it is objective whether it is conscious or not.

[5] But a detailed argument is given in Barnden (2020: Appendix section A.3), developing rougher arguments in Barnden (2014). It is a thought-experiment to the effect that, if the causal linkage within processes did not matter for their being conscious, we could string together sub-chunks of disconnected concurrent conscious processes in a multitude of different ways, such that we would be forced to conclude, absurdly, that the new processes thereby defined are consciousnesses numerically distinct from each other and from the original ones. There would be a vast multitude or even infinity of these bogus extra consciousnesses.

causal spine is essentially what makes the process a process, the claim can also be summarized as saying that *the process detects its own process-ness*. I will call this reflexivity the *auto-mattering of process-ness.*

I take it as reasonable to assume that non-conscious processes, on the other hand, do not have this *auto*-mattering quality. It may well matter *to us* that, in judging a certain configuration of matter to constitute, say, a moving car, the right internal causation is present (where that causation itself might even be projected onto the car by us, or involve human construal in some other way). But we cannot take the same line as regards what is and is not, objectively, a conscious process. A car is not a coherent process for itself, but only for us, if only by construal, whereas a conscious process is a coherent process for itself, and moreover is entirely objectively so.[6]

3 The Meta-Causal Account

3.1 A Necessary Condition for Consciousness

Given the points in section 2, MCA, while taking on board the idea that reflexivity is key in all consciousness, avoids assuming that this amounts to auto-*consciousness*. Instead, it assumes the auto-mattering of process-ness. MCA refines this notion into the notion of *pre-reflective auto-individuating auto-sensitivity* (PRAIS), and accordingly assumes that PRAIS is necessary for consciousness:-

Generic PRAIS Necessity Condition

For a process to be uninterruptedly conscious: At every moment in its time-span after its start, the process must be physically sensitive to its own causal process-ness so far (at least a little way back in time) in a way that is pre-reflective and individuates (differentiates) the process from the world outside itself.

What the condition demands is sensitivity to—some sort of being-affected-by—the causation in an (at least recent) portion of the process history, taking that causation as an entity in its own right and as a unit. The history portion *"abuts"* the current moment—i.e., it leads right up to (but does not include) the current moment.

This condition is "generic" in that it does NOT place further constraints on the nature of the sensitivity. It will be a separate, principled step to constrain the sensitivity to be meta-causal.

As a necessary condition, it is completely open about what additional conditions might be needed for consciousness. But below I'll be affirming that a

[6] An earlier footnote (no. 3) mentioned the Extra Conditions (Van Gulick 2006), needed in a HOT theory, concerning the connection between a lower-order state and a higher-order state representing it. As a generalization of our question about mattering, Van Gulick says that it is often not clear *why* these extra conditions should make a difference. Note in particular that one suggested extra condition is that the inter-state connection be causal (Gennaro 2012: 32; see also Gennaro 2016).

suitable meta-causal version of PRAIS is in fact also sufficient, indeed constitutive, of consciousness.

The auto-individuation amounts to the process's being sensitive to itself in a way in which it is not sensitive to its physical surrounds. The auto-individuation captures, partially at least, the intuition that a conscious episode is somehow sensitive to the very existence of *this particular* episode of activity, as opposed to broader activity in the world of which this episode is merely an undifferentiated part. The auto-individuation is of a weak sort, not amounting to what we could call self-identification. The process itself is not required to be able to work out what its own internal history is, before a given moment, or in any sense to find it, e.g. by applying a criterion.

I assume that, even if at each instant there is sensitivity only to a short segment of the recent causal spine (causal history), rather than to the whole previous causal spine of the process, this is enough to say that, because of the overlap of the successive segments as the process proceeds, the full process over its time-span individuates itself as a whole.

Because the auto-sensitivity is always to circumstances before the current moment, there is no absolutely strict synchronic reflexivity. But I argue in section 4.5 that we still have a "time-extended" form of reflexivity.

3.2 The Meta-Causal Version of the PRAIS Necessity Condition

Next, I argue that we should in fact adopt a more specific, meta-causal version of the Necessity Condition above:-

> ***Meta-Causal PRAIS Necessity Condition***
>
> For a process to be uninterruptedly conscious: At every moment in its time-span after its start, the process must be *meta-causally* sensitive to a part of its own causal spine that occupies some (possibly very short) time-interval abutting the current moment, and must be NOT be meta-causally sensitive to causation that is outside the process but is within the process's time-span.

The auto-sensitivity is now a matter of a *direct*, *meta-causal* influence that the mentioned prior within-process causation, taken as a unitary entity in its own right, has on the current state.

Four clarifications are needed. First, MCA does *not* take the normal view of causation as a matter of something at time A causing something at a later time B. Rather, there is a continuous stream or "spine" of causation linking infinitely many things—ordinary states or sub-streams of causation viewed as units—within the spine. Such aggregates are what "causings" are in MCA. The causation within the time-interval mentioned in the Condition is part of the process's overall spine and hence a causing in that sense. Meta-causation within a process is a spine-like aspect of the process's causal spine. It links, at least in part, such causings, but it can link these to ordinary states as well.

Secondly, at each moment in the process, the causation in the abutting interval includes only causation directly affecting state *within* that interval. It therefore does

not include the meta-causal influence on state at that moment itself. Thus, this influenceis not responding in part to itself.

Thirdly, because MCA reifies causation as a physical entity, the overall state "at" a given moment within a process includes, not just familiar types of physical state at that moment, but also causation (including meta-causation) in a time interval from that moment and going at least some way beyond it. So, as one illustration of the complex possibilities, meta-causation lying before the moment can be partly responsible for meta-causal effects on meta-causation lying soon after it.

Fourthly, just as past causation in the abutting interval meta-causally affects current state, a similar statement applies to all earlier states in the process, of course. So the causation meta-causally affecting current state includes earlier instances of that very same sort of meta-causation. Accordingly, the meta-causation affecting current state could be described as meta-meta-causation. But, of course, causation of *this* sort is present throughout the process history up to now. So we have meta-meta-meta-causation, and so on indefinitely. This might suggest a troublesome infinite regress or "ladder of meta" going upwards, especially as there are infinitely many preceding time points. We would need a ladder with an uncountable infinity of rungs! But I propose that this ladder can be viewed merely as a heuristic, descriptive matter, useful for some types of situation. Ontologically there is no progression of levels of meta-causation at all. Rather, meta-causation is always at one single level. The fact that a meta-causing interrelates some meta-causings (as well as some base-level causings perhaps) does not make it lie in a higher level than a meta-causing that interrelates only base-level causings. The former meta-causing and the meta-causings it interrelates are all at that same unique level.

Thus, altogether, the process contains a spine of causation embedded within which is a spine of meta-causation that is undifferentiated as to level.

Additionally, at the current stage of MCA's development, the Meta-Causal PRAIS Necessity Condition makes a simplifying stipulation about the auto-individuation. This is that the process is not meta-causally sensitive at all to any causation-as-such that is outside itself and occurs during the time-span of the process. This is a stark sort of auto-individuation. But I hope to loosen this in future to allow the process to have some meta-causal sensitivity to temporally overlapping causation that is outside the process, while still being sensitive in some *distinguished* meta-causal way to its own causation.

Note that MCA does allow a process to be meta-causally sensitive to causation-as-such *before* the start of that process. This is to allow sub-stretches of a continuously conscious process that start after the start of that process to be, themselves, conscious processes. Because of being part of the overall conscious process, the state at the start of such a sub-stretch must have some sensitivity to causation that lies within the process's overall stretch but before the sub-stretch.

Why should the auto-sensitivity be meta-causal? What about the alternative suggestion that the auto-sensitivity required by the Generic PRAIS Necessity Condition consists of a *representation*, within the current process state, of (a recent part of) the causal spine, or at least of the sequence of process states related to each other by that spine? The idea would be that ordinary base-level causation not only creates the current moment's state as it would otherwise do, but also includes within it a representational record of part of the causal spine. I

reject this proposal because I strongly suspect that representation cannot be made to be all of: fully objective (hence naturalizable), fully pre-reflective, and fully local. (See Egan 2020, for some discussion on this, and Shea 2018 for a detailed overview of representational approaches.) In brief: there are well-known problems of uncertainty (here, the need for a theorist's construal) about what is represented by a particular representation; complex representational structures are in danger of being reflective in involving concepts, descriptions, propositions, etc.; and teleological elements in some accounts of representation not only add further non-objectivity, but also mean that it is not a matter local to a particular physical process whether it is conscious or not. There are also the familiar problems of misrepresentation and failure to represent anything.

One way to capture the point is that it is implausible that a lower animal's states could *represent* a complex trajectory of prior internal states or causings. It is difficult to see how such a representation could avoid being complicatedly propositional in some form. One might instead propose that it could be something with structural resemblance to a partial process history. But this would bring in problems of *objectively* defining what precisely in the world the representation resembled. It might "inadvertently" resemble all sorts of things (cf. observations within the "sole-pictures" arguments in Kirk 2005: 45–53.) To make this problem worse, it is implausible that we could insist that the resemblance in representations be fully exact. Inexactness further widens the inadvertent possibilities, and weakens the possibility of objectivity of representation.

Gennaro (2012 : ch.8; 2018) says that HOT theory proponents have found evidence of higher-order states (which might be analogous to the type of representational states we are currently discussing), in some animals including some birds. See also Quilty-Dunn (2020) on perception in children and animals as arguably including discursive structure. But such evidence is far from making it plausible that all organisms that one did not wish to summarily ban from the court of consciousness are capable of the required complex representation of internal process portions.

The issue of objectivity is key in whether representation or intentionality (considering these as essentially the same thing) can be naturalized. Contrary to Gennaro's (2016) optimism that representation will be naturalized, McClelland (2020: 460 n.5) says, "Debates around the naturalization of intentionality ... have no immediate end in sight." McClelland does continue with "Nevertheless the prospects of naturalizing intentionality are better than the immediate prospects of naturalizing consciousness, so cashing out consciousness in representational terms at least improves the prospects of naturalizing consciousness." This may be true to an extent, but only if it doesn't blind one to possible non-representational cashings-out.

There are also further philosophically conceivable possibilities for the nature of the required auto-sensitivity, such as some sort of diachronic non-causal but nevertheless physical grounding. Or, perhaps the current state *directly contains* or is *partly constituted by* the prior history (rather than being merely *preceded* by it). I leave it to others to develop these proposals for our PRAIS purposes. In the concluding section of this article I mention the possibility of deploying the as-yet-inadequately defined technical notion of "acquaintance" with one's own mental states, though there with the intent of proposing that it be cashed out as meta-causation anyway.

Thus, I put forward meta-causation as the particular implementation of PRAIS as, at least, an interesting possibility for consideration.

3.3 The Constitution Conjecture and Type-Identity

I now go further and conjecture that some suitable form of meta-causal PRAIS is not only necessary for consciousness but also sufficient—in fact, *constitutive* of the process being conscious in at least a basic, pre-reflective way. I emphasize the word "conjecture" because I have weaker arguments for it than for necessity. However, I believe it is methodologically and intuitively appealing, for reasons I briefly touch on in section 7.

The statement of the conjecture involves the notion of a "slimmer" process within a given process P. The intuition is that the slimmer process is spatially within P and with the same time-interval. It can lack some of the spatial extent of P at any moment, and can lack state features that P has. As a simple and probably unrealistic illustration, if P involved the movements of certain types of molecule, the slimmer process might be missing some molecule types.

Meta-Causation/Consciousness Constitution Conjecture

Having, uninterruptedly throughout its time-span, *some suitable form of meta-causal PRAIS* (pre-reflective auto-individuating auto-sensitivity), and not containing a slimmer process that also has such PRAIS throughout, constitutes a physical process's being a [slimmest] uninterrupted conscious process. The consciousness so constituted is at least of some pre-reflective, basic phenomenal sort.

So the meta-causal account is a physicalist *type-identity* theory of consciousness. It identifies (uninterrupted) consciousness with suitable uninterrupted meta-causal PRAIS. However, unlike traditional, mind/body type-identity theories, the claimed identity is not specifically with types of brain state or other bodily state. Instead, it is with a broad type of physical state, characterized in terms of meta-causation, that could, as far as we currently know, be realized in many different sorts of physical system. It could potentially be realized in many different sorts of matter and modes of organization thereof, not necessarily confined to anything structurally similar to living matter of any sort. In short, consciousness is, potentially at least, *massively multiply-realizable*.

Of course, the physical circumstances must be such as to support the needed type of meta-causation. Notice also that a mere simulation of a conscious system, by means of a system with somewhat analogous physical structure or by means of a computer simulation program, would not be conscious unless the simulating physical system, e.g. computer, itself included the appropriate special meta-causation. But then this consciousness might have little to do with the consciousness being simulated (depending on just how the meta-causation was involved in the simulator). This point blocks a common sort of thought-experiment involving "bizarre realizations" of consciousness – cf. Van Gulick (2009).

The "some suitable form of" qualification in the Conjecture is included because it may be that not just *any* form or arrangement of meta-causal PRAIS would be

enough. In particular, I conjecture, and assume henceforth, that the auto-sensitivity at any moment includes effects on some or all of the meta-causation involved in PRAIS going forward from that moment. Further additional specifications may be needed, yet to be uncovered. Henceforth, for brevity, I will call the suitable form and arrangement through space and time of meta-causation constituting consciousness of a process *the special meta-causation [for consciousness].*

The "slimmer process" qualification is included to prevent it being the case that, for instance, a set of entirely unrelated conscious processes could be classed as a single conscious process by the account. It also prevents, say, the entire activity process in one's whole brain constituting a conscious process just because a slimmer process does so.[7]

There is an important point about the phrase "at least of some pre-reflective, basic phenomenal sort". This means that the consciousness does have a pre-reflective aspect, and may be entirely pre-reflective, but may *also* have reflective aspects. I do not yet have a specific theory of how reflective aspects are involved in consciousness, but I do assume that a conscious process in a sufficiently advanced system may include instantiation of concepts, manipulation of propositional structures, and so forth, carried by additional causation within the process. For these reflective aspects to amount to reflective *consciousness* they must be suitably yoked to the special meta-causation. I do not yet have a theory of this yoking.

Because of the identity between consciousness and a particular type of physical circumstance, we can talk about a conscious process entering into causation relations, or more briefly as (an instance of) consciousness doing so. References to instances of consciousness are just references to certain physical circumstances, even if we do not yet know the exact details of the identity.

3.4 How does a Conscious Episode Start, Keep Going, and Stop?

As well as proposing a causal *constitution* for conscious processes, MCA takes causal influences to start, sustain and terminate them.

MCA does not currently have a full account of these matters, but I assume that ordinary, base-level causation within the overall system at hand (e.g., human brain) can lead to physical states that are such that the special meta-causation constituting consciousness starts up, thereby setting up a conscious process. Causal influences from outside a conscious process can help to sustain it. Dually, ordinary causal influences from outside a conscious process can affect it in such a way that the special meta-causation can no longer continue. Of course, a conscious episode might also stop because of its internal dynamics rather than outside influences, or the stopping may be a joint result of internal and outside influences.

I also propose that the special meta-causation for consciousness has an *"auto-sustaining"* quality whereby the existence of such meta-causation up to some time-point tends to meta-cause more of the same beyond that time. Thus, while the

[7] This is not the only way of preventing undesirable consequences of these sorts, but is convenient for the present article. Also, they are only undesirable for the purposes of presenting MCA. In other contexts, there is of course nothing wrong in talking of a brain or whole person being conscious on the basis of containing a conscious process somewhere inside.

constitution of consciousness and ways in which it starts, keeps going and stops are formally separate issues, auto-sustenance is in an overlap of the issues.

3.5 What is Causation in the Account?

MCA needs there to be a fully objective and physical type of causation, one that encompasses the possibility of meta-causation. But current physics can proceed without talk of causation, and even in the philosophy of causation it is difficult to find a convincing notion of truly objective, physically real causation. (For reviews of philosophical causation theories, see Ehring 2009, Kutach 2014, Schaffer 2016.) For a start, there are overtly projective accounts, whereby causal relationships do not objectively exist but instead are merely construed by humans to exist to suit their purposes. But, on the other hand, non-projective accounts rely on such notions as counterfactuals, interventions, classes of events, or networks in which physical quantities are transferred. While I cannot argue the case here, such notions, even the last one listed, all bring in some element of construal or stipulation by people, for instance a carving up of the world or a conceptualization of it in a certain way, and therefore are not thoroughly objective. And few accounts are open to meta-causation, or have even been discussed as being open or not to it. Here the easiest non-projective case is the proposal that causation is a relation between facts, with causings just being more facts, so that meta-causal facts are straightforwardly accommodated. But to make this objective and physical would require the world to be made up of facts, ones which, moreover, are free of human construal. I find this idea difficult.

So, for the purposes of the full MCA, I have borrowed the view that there is a productive "dynamism" or "oomph" (Demarest 2017, Kutach 2014, Schaffer 2016) in the basic physical fabric of the universe, which accounts for the necessitation of later states by earlier ones. I identify causation with this dynamism. But I give the idea a radical twist by making dynamism itself a first-class citizen of the universe, with spatiotemporally located "chunks" of it able to interact dynamically with other entities in a law-governed way.[8]

I emphasize that the present article does not rely on this particular choice for the nature of physical, objective causation, except in some discussion points in sections 6 and 7. Other options for a fully physical and objective form of causation and meta-causation may come to light in the future, leading to possible new full MCAs.

[8] In Barnden (2020) I lay out my view of dynamism, including an initial mathematical formulation of how dynamism can be explicitly mentioned by an extended type of physical law, and thus appear in mathematical system equations. This appears to be the most precise regimentation of the notion of dynamism or oomph that is available.

4 From PRAIS to PRAC, and Related Matters

Here I argue that MCA plausibly implies that any conscious process possesses pre-reflective auto-consciousness (PRAC) throughout. I proceed via claims AC1–AC3 below. In these claims, for consciousness to be "at least pre-reflectively conscious [of X]" means that it *includes* pre-reflective consciousness [of X], but it may also include reflective consciousness [of X] yoked to it.

(AC1) Any conscious process is, at least *pre-reflectively*, conscious at each moment[9] of at least some aspects of *its own prior fully-internal causation*.

The pre-reflective aspect of this consciousness of causation does not, of course, involve the application of any concepts, let alone conceiving of the causation *as* causation or *as* its own. (There might in principle, in a suitable organism, be such conceiving as part of reflective consciousness of the causation. But, plausibly, such consciousness is at best rare, even in people.)

(AC2) Any conscious process is, at least pre-reflectively, conscious at each moment of *its own prior special meta-causation*.

("Special" meta-causation continues to mean meta-causation of the sort that is constitutive of consciousness.)

(AC3) Any conscious process is, at least pre-reflectively, conscious at each moment, of *its own recent ongoing consciousness.* That is, the process has PRAC throughout.

Thus, *we have derived PRAC from more basic principles,* rather than from phenomenological considerations or from the existence of reflective auto-consciousness.

4.1 The Argument for AC1

When a process is conscious of an external entity such as a red rose, the process is meta-causally sensitive to its own prior causation. As mentioned above, this "own causation" includes the causation coming in from outside, for instance *via* visual processing of the rose. So, this meta-causal influence of the incoming causation is at the root of what makes the process *conscious* of the rose in particular. In the following I will assume for simplicity of exposition that the only outer matter that the process is conscious of is the rose, but a fuller exposition would be entirely parallel. I also ignore for simplicity any outgoing causation from the process.

Now, the simplest assumption is that there is no fundamental difference of kind between

(a) the way the incoming base-level causation from the rose meta-causally influences the process

[9] Recall that this means conscious in some interval surrounding the moment.

and

(b) the way the process's other inner causation, i.e. its fully-internal causation (base-level or meta-level), meta-causally influences the process.

Moreover, the fully internal causation (the narrow causal spine) is itself directly or indirectly modulated by the incoming causation. I submit that it would be *unmotivated* to add the *extra complication* of supposing that the fact that the process is conscious *of the rose* (as opposed to something else that might have been outside) is only a matter of the meta-causal influence of type (a) and not also of the meta-causal influence of type (b). They are intimately combined. There is no reason to complicate the picture further by assuming that the process is differentially affected, through meta-causation, by the incoming causation as opposed to the fully internal causation. But then, *there is just as much reason to say that the process is (pre-reflectively) conscious of its own entirely inner causation as that it is (pre-reflectively) conscious of the rose*.

4.2 The Argument for AC2

An additional, but natural, assumption we will need is that:

> the process's special meta-causal sensitivity, at any given moment, to its prior causation includes direct sensitivity to, in particular, all the prior special meta-causation within that prior causation.

The assumption is *not* that the auto-sensitivity at the moment can in some sense *pick out* the prior special meta-causation. All that is required is that the prior special meta-causation does in fact play a direct part in meta-causally affecting the process state at that moment. Note that the prior special meta-causation is part of the process's fully internal causation.

I now appeal to a parity argument with consciousness of an external object such as a small patterned mat.

Suppose one is looking down on the mat, and, as one amongst several features of the mat's surface, there is a fine-scale colour pattern that suffuses the whole mat surface. So the patterning is a partial cause of one's conscious visual processing of the mat. I claim that this is enough to say that one is at least pre-reflectively conscious *of the patterning*. One may not separately notice the patterning, but that does not matter for present purposes. All that matters is that one's conscious state is *as directly affected by* the patterning as it is by other aspects of the mat.

Similarly, given AC1, plus the fact that the special meta-causation is fully internal, and our additional assumption above, we can say the process is at least pre-reflectively conscious of that special meta-causal aspect. The whole prior internal causation to which the process is sensitive is analogous to the mat, and the special meta-causation within it is analogous to the patterning.

4.3 The Argument for AC3

Here I appeal to the following principle, which exploits the fact that, while pre-reflective consciousness is a mental feature, it is not complicated by matters such as conceptualization and modes of presentation:

> *De-Re-ness of Pre-reflective Consciousness Statements:* Because of the lack of conceptualization in pre-reflective consciousness, if the phrase X denotes the same thing as phrase Y, then "pre-reflective consciousness of X" denotes the same thing as "pre-reflective consciousness of Y".

From the AC2 argument, we have that the conscious process at each moment is (at least) pre-reflectively conscious of the prior special meta-causation (X) that constitutes (by our identity thesis) the consciousness (Y) within the process before that moment. So, by the De-Re principle above, the process possesses pre-reflective consciousness of its own prior consciousness. Despite the prior-ness, I claim that this is enough for us to say that the process has a sort of pre-reflective *auto*-consciousness (PRAC), given that the prior consciousness is over an interval abutting (coming right up to) the current moment, as opposed to having finished some time previously. I justify the "auto" further in section 4.5.

4.4 Flavoured Auto-Consciousness

A main point in the arguments for AC1–AC3 is the intimate way in which the process's reaction to the rose is combined with its reaction to itself. This point supports the following claims:

(i) being conscious of some outer objects is just an aspect of the process's particular *auto*-consciousness on the current occasion, modulated or *"flavoured"* as it is by input from the objects; and
(ii) being auto-conscious in the particular way it is, on a given occasion of being conscious of outer objects, *is the process's way* of being conscious of the objects.

Consciousness of the outer objects and the auto-consciousness involved are two sides of the same (meta-)causal coin. *Each is a way of being the other*.

Clauses (i) and (ii) have some resonance with Strawson's (2017: 60) statement that "the contentual features of states of awareness—more precisely, the contentual features of states of awareness that involve content other than whatever content is involved in simple awareness of awareness— should rather be seen as modifications of awareness." (However, a clearer parallel would arise if that last word were replaced by "awareness of awareness".)

The clauses are also in the spirit of the adverbial-style rendition by Rowlands (2016) of Sartre's view of PRAC. One difference of emphasis is that his statement there (p.117) that "I am [pre-reflectively conscious] of my [reflective] acts of

consciousness [of some outer objects] because of the way in which those acts present [the objects] to me" is analogous to (ii), and he presents no analogy to (i). So MCA has an entirely fair coin, whereas it seems Rowlands's coin might always land one way up.

There's a sense in which auto-consciousness is more fundamental than outer-directed consciousness. In MCA, there can in principle be an entirely isolated conscious process, with no outer-directed consciousness. But you can't have outer-directed consciousness without auto-consciousness.

4.5 The Time-Extended Reflexivity of Meta-Causation and Consciousness in MCA

The meta-causation constituting the auto-sensitivity, at some moment within a process, to its prior causation is not reflexive in a strict sense, as it doesn't take itself as a relatum. However, if we take the "long view" over the time span of a conscious process, it is reasonable to say that the process and its meta-causation, *as time-extended matters,* are sensitive to themselves, and are so *continuously and throughout,* in a thoroughly *internal* way, for the following reasons.

The state going forward from a given moment in the process, including special meta-causation going forward from there, is meta-causally affected, across the moment, by the meta-causation in a part of the process's causal spine *abutting* that moment. Let's label that meta-causation across the moment as the "linking" meta-causation. Recall that the causation in the prior spine part acts *as a unit* on the cause side of the linking meta-causation. So there is no temporal gap between the cause side of the linking meta-causation and the effects going forward from the moment, albeit that both the cause side and the effects are time-extended, and the linking meta-causation is itself essentially a time-spanning matter, as all causation is. I contend that, considering now all the successive moments within the process, the gapless linking at each moment warrants the statement that the process, including its special meta-causation, has *time-extended reflexivity continuously throughout*.

Indeed, the same argument applies to any sub-stretch of the process (a sub-process occupying any given sub-span of the time-span of the process). So, it's not just that the whole process has a global property of time-extended reflexivity. This property applies to each sub-stretch no matter how finely we divide the original process.[10]

Furthermore, the time-extended reflexivity of the meta-causation is *"internal,"* in the sense that all the meta-causation involved in making the overall meta-causation reflexive is, itself, within that overall meta-causation. It's not that there is meta-causation *outside* the meta-causation of the process or sub-process, relating that meta-causation to itself. In a time-extended way, the (sub-)process's meta-causation internally relates itself to itself.

[10] I am putting aside here any practical limits there may be in a given sort of physical substrate on how short a time-span can be and still support meta-causation. I don't know if there are such limits, but even if there are I believe it only makes a minor difference to the reflexivity claims I am making.

Parallel reasons justify our saying that a conscious process, as a time-extended entity, is genuinely *auto*-conscious in a time-extended way, given the nature of the arguments leading to AC3 above. Within a conscious process, at each moment the consciousness going forward from that moment is conscious of the consciousness going back at least a little way before the moment, with no time gap in between, albeit that both conscious sub-episodes just mentioned are time-extended and that one is after the other.

And the reflexivity here is internal to the process's consciousness itself, in a sense analogous to the above internality. The connection between the conscious sub-episodes is meta-causation spanning that moment, and *such connectivity is itself part of the special meta-causal whirl forming the consciousness over any given interval spanning that moment*. Recall from section 2.2 that consciousness is not defined at instants, so we must not think of the connection between the conscious sub-episodes as instantaneous and lying merely *on* the border between them. Thus, the across-moment linkages serving the time-extended auto-consciousness of the process are themselves an intrinsic part of the process's consciousness.

A partial analogy with an existing account of consciousness may help. In Gennaro's model (WIV, or Wide Intrinsicality View, Gennaro 2012, 2016), a conscious state includes a "metapsychological" component (MET) that bears a representational relation to another component, M, which may for instance be a first-order perceptual representation of an external scene. Gennaro (2016) indicates that the fact that no component of the conscious state literally refers to itself should not prevent our saying that, as a whole state, it refers to or represents itself. The diachronic relationship in MCA between the meta-causation following a moment and the meta-causation preceding it is a rough diachronic analogue of the synchronic separation of the two mentioned WIV components, MET and M. But I would say that MCA's diachronic relationship is a less striking separation than WIV's synchronic cleft.

5 What's It Like?

5.1 Strawson on Intrinsic Self-Intimation and (Non-)Loopiness

This article's approach has a strong affinity with Galen Strawson's theorizing about the long-standing notion of *self-intimation* (Strawson 2017: ch.8; see also McClelland 2020: 462/3). The idea is that self-intimation is self-consciousness intrinsic to experience, rather than being an extra consciousness (separate component of one's conscious state). (See also Montague 2017, Textor 2015.) Strawson says that consciousness "comports" consciousness of itself, where "to comport" means to contain wholly within itself. The reflexivity of consciousness is part of what actualizes consciousness from moment to moment, as part of its "dynamic essence"—there is an *"infinite whirl"* in Strawson's words.[11]

[11] On terminology:- (a) In summarizing some of Strawson's claims, I have translated "awareness" as "consciousness" for consistency with the present article. (b) Strawson's "whirl" has inspired my use of "whirl" below.

Similar claims emanate from MCA. Crucially, having phenomenality at all is equated with having a suitable pattern of internally reflexive meta-causation, in the time-extended sense explained in section 4.5. This pattern can be modulated or flavoured, so to speak, because of the particular external objects causally impinging on the conscious process. Being *conscious* of an external object just is, intrinsically, to have *this thus-adjusted pattern of internally reflexive meta-causation*. But having this pattern is, just as much, being conscious of this very meta-causation, by the arguments in section 4. So the process is intrinsically and inseparably both conscious of the external object and conscious of its own consciousness. The consciousness comports consciousness of itself, because it just is consciousness of both itself and the external object. It is only a matter of our theoretical analysis that might make it seem that there are two separate components of consciousness that have some difficult connection.

Further, in MCA, we see that there is a sense in which the reflexivity of consciousness is what "actualizes" consciousness from moment to moment, as claimed by Strawson. But we can clarify the perhaps puzzling circularity of this statement, by means of our meta-causal unpacking of consciousness. It is the special, reflexive *meta-causation* that actualizes consciousness from moment to moment. That statement is now non-circularly expressing the point that consciousness just is a matter of the reflexive meta-causation, equipped with its auto-sustaining quality as described in section 3.4. But that reflexivity is foundational to the consciousness. We can take Strawson's "actualizes" to be a combination of "realizes" and "sustains the proceeding of". The particular reflexive meta-causation just is the consciousness, but by virtue of that (time-extended) reflexivity it sustains itself through time. This article's proposal is thus very much about PRAC itself being the "dynamic essence" of consciousness and being an "infinite whirl" of meta-causation.

Strawson (2017: ch.8) says that "The metaphysical complexity of self-intimational complexity is part of what the existence of [the] phenomenological simplicity consists in." Here he is alluding to a tension he perceives between the phenomenal "flatness," or "non-loopy" quality, of conscious episodes and the actual, metaphysical reflexivity (loopiness) of consciousness inherent in its being self-intimating. He describes the flatness in saying "everything that is experienced, however multimodal, is on a single experiential plane, the only experiential plane there is, which is, quite simply, the plane or 'field' of experience."

Our meta-causal analysis allows clarification of Strawson's intriguing point that the phenomenal *simplicity* consists in the metaphysical *complexity*. The opportunity for clarification arises because Strawson's metaphysics itself is ultimately phenomenology-based, whereas MCA's meta-causal metaphysics is independently based. In MCA, we have the loopy complexity, i.e. internal reflexivity, of the special meta-causation of a given conscious process. The process's phenomenality consists in the process having this loopily-complex meta-causation. The latter loopiness *shows up holistically simply as phenomenality itself*. The whole loopiness is required in order for there to be phenomenality *at all*. There is no way of dissecting the meta-causal loopiness into layers that individually have their own separate phenomenological nature.

This does not mean that we cannot at least roughly dissect the loopy meta-causation along other dimensions that cut across it. Suppose a conscious process is

conscious of various external objects. The overall causation, including meta-causation, within a conscious process must have, in *some* sense, different aspects that correspond to the different external objects; and it may be that these aspects are roughly separable from each other, i.e. there are different (if overlapping or interacting) sub-streams of causation corresponding to the different objects. Each sub-stream would have loopiness that forms an aspect of the overall loopiness of the conscious process.[12]

5.2 Immediacy and Non-Relationality

Frank (2007, 2016, 2019) champions various crucial, interrelated claims, linking also to self-intimation. They include "Fichte's original insight" that reflective self-consciousness cannot exist unless there is already in place an immediate non-reflective reflexivity of some sort, whereby the subject of consciousness already has some sort of immediate access to itself. "Immediate" (alternatively "immediated", "unmediated", etc.) apparently means being in some sense seamlessly joined and, in particular, not mediated by such things as conceptualization, self-identification by description, or inference. Frank also wants the reflexivity of consciousness to be non-relational—not to be analysed as the subject bearing a relation to itself, but rather as a unary property of the subject.

It is beyond the scope of this article to discuss "subjects" except in the very "thin" sense of subject that Strawson (2017: chs 3, 9, 10, 11) amongst others has discussed. This article is most attuned to the extreme version of this, whereby the subject of an episode of consciousness is that episode itself. Thus, I will here interpret the above claims as being about the immediacy of the episode's access to itself and the non-relationality of its auto-consciousness.

MCA has a somewhat mixed relationship to Frank's requirements. A first point about immediacy is that the everyday temporal meaning of "immediate" is relevant. If there were mediation as above, one might expect temporal non-immediacy, and similarly vice versa. Accordingly, Frank (2019) is critical of the time lag, however minute, admitted for instance by Horgan & Kriegel (2007) to be needed for self-representational states to arise, in their approach. While MCA is based on meta-causal linkage, which is inherently a time-spanning matter, it is the causation in an interval *abutting* a given time T within a process that meta-causally affects state at T (including meta-causation a little way onwards from T). Thus, this meta-causal linkage involves no temporal gap between the end of that interval and T.[13] There is a weak, partial form of temporal immediacy here. But there is also very strong immediacy of a non-temporal sort, because the meta-causation does not go through any intermediary. That is in actuality a main reason for MCA's introduction of it.

[12] But the sub-streams must not be separate enough to count as [slimmest] conscious processes in their own right, because of the "minimal slimness" provision (section 3.3).

[13] The interval is at least "open" on the right, in the mathematical sense: it contains time points arbitrarily close to T. T is a limiting point for the points in the interval, and could be said to be the "end" of the interval, but is not in the interval.

While the reflexivity of meta-causation is only indirectly related to the reflexivity of consciousness, as mentioned in section 5.1, the non-mediation involved in the former reflexivity could be argued to supply a sort of non-mediation in the latter. However, because consciousness is a complex whirl of meta-causation being sensitive to itself and other causation, rather than just being based on, say, a single meta-causal link, I am not sure the non-mediation would be strong enough to fit what Frank would like to see.

I turn now to non-relationality. Causation and meta-causation in MCA can be viewed as relations, of a rather unusual, infinitary sort, in that any [meta-]causing links the states at infinitely many time points forming some interval. Hence, consciousness has an intensely relational basis. However, since, as mentioned in section 5.1, phenomenality consists holistically of the whole whirl of special, reflexive meta-causation in a conscious process, rather than being structured as the meta-causation is, perhaps the reflexivity of consciousness could be analysed as a unary property. I pose this as an issue for future consideration.

Finally, Frank (2016, 2019) is concerned that the subject should *know* the subject-object identity in immediate self-consciousness. The topic of knowledge is way beyond the confines of the present article, but I should mention that consciousness in MCA is meta-causally affected by the meta-causation within itself that serves the role of making the consciousness reflexive. In that sense, consciousness has an immediate acquaintance with that which constitutes its own reflexivity. (I'm not here seeking to use the term "acquaintance" in any technical sense. See section 7 for a brief word on technical acquaintance.)

5.3 For-Itself-Ness and For-Me-Ness in the Account

MCA readily enables a particular explication of the for-itself-ness and hence for-me-ness of consciousness, as follows.

Consider Sally being conscious of a rose. The rose-based phenomenology is, of course, not accurately described as being based solely on the rose. Rather, as Kriegel (2009) and others have pointed out, Sally's phenomenology is not just a matter of redness of the rose, roundness of the rose, certain structural features of the rose, etc. etc., but also somehow, perhaps very subtly and implicitly, involves phenomenology of for-herself-ness. This is not separate phenomenology, in that it is integrated with the redness etc.: the phenomenology is that of redness-for-herself-ness. (For-herselfness would usually be talked of as for-me-ness and mine-ness, but I find the third-person flavour healthy here. See Guillot 2017 and Zahavi 2018 on the range of different notions available in the vicinity of for-me-ness and mine-ness.)

My view here is first that this for-herself-ness feeling is actually a perceptually/cognitively-*affecting*-herself-ness feeling, which I will now just abbreviate to an affecting-herself-ness feeling.[14] So Sally's rose-based phenomenology contains phenomenology of redness-affecting-herself, roundness-affecting-herself, silky-texture-affecting-herself, etc., and of inter-relatedness-[of different aspects of the rose

[14] Caution: I do not mean "affecting" in the sense of giving rise to affect—i.e., emotion, evaluation, mood, etc.— although it could often give rise to it.

with each other]-affecting-herself. The rose-based phenomenology is *entirely* of a for-herself-ness type, in numerous varied subtypes. This appears to conform to, for instance, Zahavi's (2018) position that for-me-ness isn't some specific feeling, but instead the first-personal presence of all "my" experiential content.

In MCA, the affecting involved here is actually a matter of causation from the rose into the conscious process, combined with the thereby-modulated meta-causation within the process. (But as always, even if the consciousness involves reflection, we are not assuming use by Sally of *concepts of* meta-causation.) The for-herself-ness is just an aspect of the already argued point that consciousness of an external object and consciousness of own consciousness on that occasion are two sides of the same causal and meta-causal coin (see section 4.4). The meta-causal account provides an account of qualities such as for-herself-ness that does not rest circularly on talk of consciousness.

In making these points, I am downplaying the more "personal" aspects of "me" or "herself", and concentrating on the for-herself-ness being at core a non-personal for-itself-ness, the "it" being the conscious process. However, I envisage that, to the extent that consciousness can involve a phenomenology of a personal "oneself", this would be an enrichment of the for-itself-ness phenomenology, based on an enrichment of the (meta-)causal flux involved in the latter. The yoking to for-itself-ness of the extra element that is needed for personal herself-ness here may be similar in quality to the envisaged yoking of reflection to PRAC. Moreover, reflectiveness and that extra element may overlap, but the issue is beyond the scope of this article.

5.4 The Appearance of Transparency

As already noted, for instance in section 5.1, the meta-causal account supports the "intrinsicality" thesis that consciousness of some external object intrinsically includes at least pre-reflective consciousness of (aspects of) that consciousness. Indeed, the core of MCA's argument here was that, given the structure of (meta-) causation involved in the episode, it is difficult to draw a line between the outer-directed and inner-directed aspects of the consciousness. As part of this point, these aspects are on a par with each other, in the sense that the consciousness is *just as much* auto-consciousness as it is outer-directed consciousness. And actually there is even a case for it being *more* a case of auto-consciousness than outer-directed consciousness, as more-direct and more-local causality is involved, and as auto-consciousness is more fundamental than outer-directed consciousness (section 4.4).

But then there is tension with the understandable common-sense intuition that the experience, where for instance Sally is consciously seeing an external rose under normal circumstances and not reflecting on this situation, is one where there is merely phenomenology of a *rose* having certain qualities (and, of course, of surrounding or background objects having certain qualities, but I will ignore such objects for simplicity). So one's consciousness is allegedly "transparent" in the metaphorical sense that it's like an invisible pane of glass through which one is seeing just the rose, or like one's eye. By contrast, from the parity mentioned a moment ago (of the inner and outer-directed consciousness) one might expect (a) there also to be

phenomenology of the consciousness itself having certain qualities (alongside the rose-based phenomenology) or (b) there to be an integrated phenomenology of the rose-together-with-this-consciousness having certain qualities.

Of course, the for-herself-ness point above already diffuses the tension to an extent, but one might still understandably claim that the phenomenology is intuitively more strongly rose-focused than focused on the person or her consciousness. This apparent rose-bias can be explained as follows.

First, there is the simple point that, in human beings, the outer-directed aspect of the experience is probably somewhat reflective (e.g., concept-involving) whereas the inner-directed aspect is probably not, under normal circumstances. The outer-directed aspect of consciousness is simply richer.

Secondly, a point specific to MCA is that while there is some sort of parity between the way the incoming causation from the rose works and the way the fully internal meta-causation works, it remains the case that it's the incoming causation that is at the *root* of the current, *particular,* "rose-modulated" form of the internally-reflexive meta-causation lying within the process, and is the cause of its distinctiveness compared to what would arise in other experiences. There remains an inherent asymmetry in the causal involvement of the rose and causal involvement of the causation within the process itself, and this could be enough to explain the "rose bias" in the phenomenology.

Note that MCA does affirm that the "transparency" is merely apparent and doesn't survive technical analysis. The "[mere] phenomenology of a rose" mentioned early in this subsection just is auto-consciousness as well as rose-directed consciousness. Saying that Sally just consciously-sees the rose and is not conscious of her own consciousness is merely a heuristic, commonsensical, inaccurate description that hides the reality that consciously-seeing *is* being conscious of her consciousness in a particular, rose-flavoured way.[15] (I continue to use "flavoured" metaphorically, of course!)

6 Further Discussion: Panpsychism and Bathypsychism

As I have drawn strong links between MCA and Strawson's work on self-intimation (section 5.1), I should mention that MCA doesn't share Strawson's support for panpsychism (e.g., Strawson 2016), whereby consciousness (in some primitive form) constitutes the bottom-most level of the universe, with physical matter etc. somehow based on it. One implication of such panpsychism is that (primitive) consciousness is spatiotemporally ubiquitous.

MCA does not claim that consciousness is spatiotemporally ubiquitous, let alone that everything is built out of consciousness. I stress this because MCA does allow the *possibility* that *meta-causation in general* is ubiquitous or near-ubiquitous,

[15] In these brief remarks I cannot further defend the view expressed, or engage in detail with the long-standing, large and diversely-opinioned literature on transparency. For an introduction, see Gennaro (2018: 108) and material in several chapters in Kriegel (2020).

without committing to its actually being so. There might be, in principle, so-far-undetected non-zero meta-causal influences all over the place. At the least, they might not be confined to animate beings or even to particularly complex systems. The point is, however, that consciousness consists of very special meta-causation, one feature of which is that it is internally reflexive in a time-extended way. So a given degree of spatiotemporal prevalence of meta-causation in general does not imply anything like the same degree of prevalence of consciousness.

Spatiotemporal ubiquity is a sort of "horizontal" matter, whereas the question of how far "down" consciousness goes in the physical fabric of the universe is a "vertical" matter. Here Strawson's panpsychism has more contact with MCA. Now, the present article has been at pains *not* to specify the level at which meta-causation lives in the physical universe (recall section 3.5). It could even be a high-level matter, explicable in perhaps currently-unfamiliar low-level physical terms. However, recall from section 3.5 that the full MCA claims that the meta-causation is meta-oomph (meta-dynamism), where oomph in general is at the very bottom of the universe. Hence, it is *possible* that the special meta-causation and thus consciousness could exist, here and there, in simple systems definable at a very low level in the world. It may not be confined to systems that have high-level structure, such as living organisms.

In sum, MCA is potentially *"bathypsychic,"* where *"bathy"* alludes to depth within the physical realm (cf. "bathysphere" submersibles). Strawson-style panpsychism subscribes both to ubiquity and actual, extreme bathypsychism, intimately combining them, while MCA is *open* both to bathypsychism and (less enthusiastically) to ubiquity. However, MCA, whether in its full version or not, *claims* neither that there is bathypsychic consciousness nor that consciousness is spatiotemporally ubiquitous.

Strawson (2016) says that causation is "nothing over and above the existence of energy" and "the causal laws for our universe describe … the behavioral form of energy … ." This may be an appealing view, although I am unclear what the behavioural form of energy is if not just energy *tout court* (perhaps it's oomph??). But I would still argue for a meta-level: a meta-energy distinct from energy as we know it, with consciousness requiring some special configuration of meta-energy. Strawson admits that we don't know how energy could be experiential. A "meta-energetic MCA" in place of the current meta-oomphatic MCA could conceivably help here. It could make consciousness bathypsychic, but it would still not make it basic to everything else. Meta-energy in general, as opposed to a special configuration of meta-energy required for consciousness, would be more fundamental.

7 Concluding Remarks

I hope to have presented some grounds for supposing that, if MCA—the meta-causal account of consciousness—is at all reasonable, it provides a fruitful basis for further justification and exploration of PRSC/PRAC (pre-reflective self/auto -consciousness). In particular, we have seen how one can plausibly infer that consciousness includes PRSC from a hypothesis that it includes a new type of auto-sensitivity—sensitivity to own causal nature—that is not itself defined in terms of

consciousness. This auto-sensitivity is also conveniently dubbed as auto-mattering of process-ness. Further work could continue on issues I have discussed, such as: the intrinsic involvement of auto-consciousness in consciousness, with outer-directed consciousness being just a flavoured auto-consciousness; forms of immediacy and non-relationality of consciousness; for-me-ness; and transparency intuitions. There are other prominent issues that MCA may illuminate, including for instance the possibility that a technical notion of acquaintance with one's own mental states and processes, which has been proposed as a basis of consciousness (see, for instance, the critical discussion in Gennaro 2016; and essays in Knowles & Raleigh 2019), could be partly a matter of meta-causation. Such explorations will, hopefully, not only advance the philosophy surrounding PRSC but also turn around to act as evidence that the meta-causal account itself is on the right track.

In fact, it was assumptions of processual auto-sensitivity, full physical objectivity of consciousness, and the need to avoid complex reflective elements such as representation of process portions and usage of concepts, that led to the idea of meta-causation as an aspect of the physical world. This then required an objective, physical form of causation in general, and prompted the adoption (in the full MCA) of a basic-level physical dynamism in the world as that form. All in all, a consideration of the intimate reflexivity of consciousness suggested a radical new view concerning physical causation on the one hand and a radically new detailed mechanistic framework in which to couch consciousness on the other.

Meta-causation is certainly a radical suggestion metaphysically, though it is not unprecedented, and it is even more radical, and apparently completely unprecedented, as an addition to physics. It is also highly weird intuitively, especially in the particular form of MCA's intricate, internally reflexive meta-causation, with no level distinctions, as discussed in sections 3.2 and 4.5. I see this weirdness as a feature, not a bug, because of the weirdness of consciousness itself in comparison to what we take to be the physical world. I am not just arbitrarily conflating two weirdnesses. Rather, there is a principled similarity between the two that means they should at least get acquainted even if marriage is premature. The special internally reflexive meta-causation has just the sort of self-interiority that we sense consciousness to have, and which makes consciousness so mind-bending to think about. The self-interiority of the continuously-operating internally reflexive meta-causation, though time-extended, is much tighter than the reflexivity in existing attempts to make complex, articulated subsystems include self-reference. MCA provides us with a sort of self-interiority that is tightly wound on itself in a deeper sense. In the full MCA at least, the self-interiority is all the way down (fully "bathy"), whereas self-reference in the articulated subsystems proposed in many current consciousness theories is something one ceases to see as one drills down just a little into the mechanistic innards.

Relatedly, I see profound connections between MCA and ideas from certain past researchers, notably Fichte (1982), notwithstanding his idealism *versus* MCA's physicalist realism. I have only recently started to look at these connections, but the basic point, in the Fichte case, is the idea that at the basis of everything is action that intrinsically acts upon itself and is its own deed (though some of his ideas on consciousness and causation conflict with MCA). Preliminary parallels

and non-parallels to aspects of Fichte (1982) are expressed in Barnden (2020: section 5.4.4). They are tentative, given that Fichte is notoriously difficult to interpret (as emphasized by the translators' Preface in Fichte 1982), but the importance, for current consciousness theory, of listening to Fichte and related researchers is extolled by Frank (2004, 2019).

I am tempted to propose that a certain extreme, pure version of the reflexive whirl of meta-causation is able to exist, at least in principle: *viz*, as an isolated pure consciousness, with no base-level causation involved at all. There is just some configuration of meta-causation acting upon itself, and indeed where the meta-causal self-acting *is* itself. An intriguing connection on this point is to the following claim by Kierkegaard:[16]

> *"The self is a relation that relates itself to itself or is the relation's relating itself to itself in the relation; the self is not the relation …, but the relation's relating itself to itself."*

To judge by Frank's (2004) discussion, what Kierkegaard meant here may be very different from the way I am tempted to take it. For what it is worth, if we read "the self" as alluding to consciousness in general or to a person's conscious episode on a given occasion, as appropriate, and, for sheer curiosity, specify the mentioned relation to be that of meta-causation, we get the following, with a little grammatical massaging:

> *"Consciousness is meta-causing that exerts meta-causation between itself and itself, and a conscious episode is some meta-causation's having meta-causal influence upon itself, that influence being part of that very meta-causation; consciousness is not the relation of meta-causation as such, but rather the meta-causal relating of meta-causing to itself."*

Acknowledgments The article has benefitted from constructive suggestions by the special issue editors and the anonymous reviewers.

Availability of data and material N/A

Author's contribution 100% in all respects. (Sole author.)

Declarations

Conflicts of interest / Competing interests none of any type relevant to this article.

Code availability N/A

[16] The quotation was first brought to my attention, in another translation, by Jeff Yoshimi (p.c., 2019). The displayed form is from Frank (2004).

References

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

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

Demarest, Heather. 2017. Powerful properties, powerless laws. In *Causal Powers*, ed. Jonathan Jacobs, Chapter 4. Oxford Scholarship Online.

Egan, Frances. 2020. A deflationary account of mental representation. In *What are Mental Representations?* ed. Joulia Smortchkova, Krzysztof Dolega, and Tobias Schlicht, 26–53. Oxford University Press.

Ehring, Douglas. 2009. Causal relata. In *The Oxford Handbook of Causation*, ed. Helen Beebee, Christopher Hitchcock, and Peter Menzies, 387–413. Oxford University Press.

Ellis, Brian. 2013. The power of agency. In *Powers and Capacities in Philosophy: The New Aristotelianism*, ed. Ruth Groff and John Greco, 186–206. Routledge.

Fichte, Johann G. 1982. *The Science of Knowledge: with the First and Second Introductions. Edited and translated by Peter Heath and John Lachs*. Cambridge University Press.

Frank, Manfred. 2004. Fragments of a history of the theory of self-consciousness from Kant to Kierkegaard. *Critical Horizons* 5 (1): 53–136.

Frank, Manfred. 2007. Non-objectal subjectivity. *J. Consciousness Studies* 14 (5–6): 152–173.

Frank, Manfred. 2016. Why should we think that self-consciousness is non-reflective? In Miguens *et al* (2016), 329–353.

Frank, Manfred. 2019. From "Fichte's Original Insight" to a moderate defence of self-representationalism. *ProtoSociology* 36: 36–78.

Gennaro, Rocco J. 2012. *The consciousness paradox: consciousness, concepts and higher-order thoughts*. MIT Press.

Gennaro, Rocco J. 2016. The "of" of intentionality and the "of" of acquaintance. In Miguens *et al* (2016), 317–341.

Gennaro, Rocco J. 2018. Representational theories of consciousness. In *The Routledge Handbook of Consciousness,* ed. Rocco Gennaro, 107–121 . New York / London: Routledge.

Guillot, Marie. 2017. *I me mine:* on a confusion concerning the subjective character of experience. *Review of Philosophy and Psychology* 8: 23–53.

Horgan, Terry, and Uriah Kriegel. 2007. Phenomenal epistemology: what is consciousness that we may know it so well? *Philosophical Issues* 17: 123–144.

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

Knowles, Jonathan, and Thomas Raleigh, eds. 2019. *Acquaintance: new essays*. Oxford University Press.

Koons, Robert C. 1998. Teleology as higher-order causation: a situation-theoretic account. *Minds and Machines* 8: 559–585.

Kovacs, David M. 2019. The question of meta-causation. In *Proceedings of the FraMEPhys/Gothenburg Conference on Metaphysical Explanation in Science*, Birmingham, UK, 10–11 January 2019.

Kovacs, David M. 2021. The question of iterated causation. *Philosophy and Phenomenological Research*. https://doi.org/10.1111/phpr.12782.

Kriegel, Uriah. 2009. *Subjective consciousness: a self-representational theory*. Oxford University Press.

Kriegel, Uriah, ed. 2020. *The Oxford handbook of the philosophy of consciousness*. Oxford University Press.

Kutach, Douglas. 2014. *Causation*. Polity Press.

McClelland, Tom. 2020. Self-representationalist theories of consciousness. *In Kriegel* 2020: 458–481.

Miguens, Sofia, Gerhard Preyer, and Clara Bravo Morando (Eds). 2016. *Pre-reflective consciousness: Sartre and contemporary philosophy of mind.* Routledge. (Citations to chapters rest on the 2019 paperback version.)

Montague, Michelle. 2017. What kind of awareness is awareness of awareness? *Grazer Philosophische Studien* 94: 359–380.

Oizumi, Masafumi, Larissa Albantakis, and Giulio Tononi. 2014. From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology* 10 (5): e1003588.

Quilty-Dunn, Jake. 2020. Concepts and predication from perception to cognition. *Philosophical Issues* 30: 273–292.

Rowlands, Mark. 2016. Sartre on pre-reflective consciousness: the adverbial interpretation. In Miguens *et al* (2016), 101–119.

Schaffer, Jonathan. 2016. The metaphysics of causation. . In *The Stanford Encyclopedia of Philosophy*, Fall 2016 edition, ed. Edward N. Zalta. http://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/ accessed 1 December 2016.

Schechter, Elizabeth. 2018. The unity of consciousness. In *The Routledge Handbook of Consciousness*, ed. Rocco Gennaro, 366–378. Routledge.

Seibt, Johanna. 2013. Process philosophy. In *The Stanford Encyclopedia of Philosophy*, Fall 2013 edition, ed. Edward N. Zalta. http://plato.stanford.edu/archives/fall2013/entries/process-philosophy/ accessed 1 March 2014.

Shea, Nicholas. 2018. *Representation in cognitive science.* Oxford University Press.

Strawson, Galen. 2016. Mind and being: the primacy of panpsychism. In *Panpsychism: contemporary perspectives*, ed. Godehard Brüntrup and Ludwig Jaskolla (Eds), Oxford Scholarship Online, October 2016.

Strawson, Galen. 2017. *The subject of experience*. Oxford Scholarship Online, March 2017.

Textor, Mark. 2015. "Inner perception can never become inner observation": Brentano on awareness and observation. *Philosophers' Imprint* 15 (10): 1–19.

Van Gulick, Robert. 2006. Mirror mirror — is that all? In *Self-Representational Approaches to Consciousness,* ed. Kenneth Williford and Uriah Kriegel, 11–39. Cambridge, UK: Cambridge University Press.

Van Gulick, Robert. 2009. Functionalism. In *The Oxford Handbook of Philosophy of Mind*, ed. Brian P. McLaughlin, Ansgar Beckermann, and Sven Walter, 128–151. Oxford University Press.

Van Gulick, Robert. 2010–2011. Non-reductive physicalism and the teleo-pragmatic theory of mind. *Philosophia Naturalis* 47–48 (1–2): 103–123.

Zahavi, Dan. 2018. Consciousness, self-consciousness, selfhood: a reply to some critics. *Review of Philosophy and Psychology* 9 (3): 703–718.