*(please do not cite or distribute without permission)*

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

John A. Barnden

Professor Emeritus of AI, School of Computer Science
and
Research Associate, FraMEPhys Group, Dept of Philosophy

University of Birmingham,
Birmingham B15 2TT, UK

jabarnden@btinternet.com
ORCID: 0000-0001-8062-2288

**Abstract**

I present considerations surrounding pre-reflective self-consciousness (PRSC) arising in work towards a new physicalist, process-based account of [phenomenal] consciousness. The account is called *the meta-causal account* because it identifies consciousness with a certain type of arrangement of meta-causation. Meta-causation here is where an instance of causation is itself a causal relatum. The proposed type of arrangement involves a sort of time-spanning 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. 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. 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. These ideas also naturally provide a specific account of for-me-ness in meta-causal terms, an explanation of certain ways in which consciousness might be taken to be transparent, and enriches discussion of the unity of consciousness. The article briefly comments on how reflective consciousness might be yoked together with pre-reflective consciousness. The account is non-egological at base but allows for egological enrichment in relatively advanced forms of consciousness.

**Keywords:** pre-reflective self-consciousness; meta-causation; physicalism; theories of consciousness.

# 1 Introduction

In this article I take a physicalist stance about (phenomenal) consciousness, and assume that consciousness is an objectively existing aspect of the physical world. I find claims that all consciousness intrinsically involves pre-reflective self-consciousness (PRSC) persuasive, and wish to have a physicalist account of consciousness and especially PRSC. This article in part presents such an account in summarized form, and is thus part of a stab at solving the hard problem of consciousness, but it more pointedly tries to provide evidence that the account helps to provide additional and firmer rationale and justification for certain discussed aspects of consciousness and especially PRSC, and that the account provides a fruitful basis for further exploration of consciousness and particularly PRSC.

The account is here called *the meta-casual account.* Meta-causation in my sense (al so called higher-order causation) is where instances of causation can themselves serve as causal relata. This is all within a physicalist framework. As will be explained in more detail below, the account has it that a physical process is (uninterruptedly) conscious if and only if throughout its existence it possesses a suitable form of meta-causally realized *pre-reflective auto-individuating auto-sensitivity* or PRAIS (where the notion of sensitivity used is not itself defined in terms of consciousness). The central notion is that at any moment, prior causation within the process meta-causally affects the course of the process going forward in a distinctive way. This leads to a sense in which consciousness is at heart a matter of *internally reflexive meta-causation* (meta-causation that meta-causally affects itself, that self-affecting itself being within that meta-causation).[1] The brunt of the present article is to argue for various plausible consequences and ramifications of the account, which are briefly as follows.

One consequence is that, given the account's meta-causal explication of consciousness, a conscious process will automatically possess pre-reflective self-consciousness (PRSC), albeit perhaps with only the thinnest notion of self. The significance of this derivation of PRSC is that, while PRSC may seem to many researchers (including myself) intuitively to be present in consciousness and thus might be just adopted as a premise or argued for on purely phenomenological grounds, it is advantageous for us to have an independent argument for PRSC that arise naturally out of naturalistic, non-phenomenological explication of consciousness.

Secondly, the account provides an explication of the notion of "self-intimation" as discussed in Strawson (2017), which for me is one of the more attractive discussions surrounding PRSC (see also Montague, 2017), and the related notion of PRSC being thoroughly intrinsic to consciousness (as in those works and Textor 2015, following Brentano).

Thirdly, the account provides a understanding of the "for-me-ness" or more generally the "for-itself-ness" of consciousness at a phenomenological level yoked to a particular meta-causal explication of this notion.

Fourthly, the account readily allows one to address concerns about the "transparency" of perceptual consciousness.

[1] I hope it is obvious that what is crucial here is causal structure *inside* the process. What is NOT going on is any attempt to define what an instance of consciousness is by any causal or other functional role it might play with *other* matters in the world. Causal relationships with the world outside the consciousness instance are operative in what the consciousness "perceives" and what external "actions" it causes, but not in the very question of what it is for the process at hand to be a conscious one.

Finally, the account provides a detailed basis for exploring the possible unity of consciousness (Masrour 2020, Schechter 2018). It finds taking a Brentanean view of unity argued for by Textor (2015) to be somewhat congenial but argues for a different view.

The account itself (as opposed to those consequences and ramifications, which are new to this article) has been presented and argued in detail in a recent journal publication, but had an earlier, sketchier form in a 2014 publication. The recent work includes a mathematical framework for fitting the account into physical laws and system equations, as well as the philosophical ideas, which are what the present article is mainly concerned with. The present article can do no more than briefly present the motivations, assumptions and claims of the account, together with some indication of the justificatory arguments.

The paper is structured as follows. Section 2 summarizes the meta-causal account of consciousness. Section 3 presents the argument for the first consequence above, that PRSC follows naturally from the account, given the pre-reflectiveness and reflexivity of PRAIS. Section 4 presents the remaining consequences and ramifications mentioned above. Section 5 concludes and mentions lines of ongoing and future research on the account.

But first, I include here some important terminological notes.

I will always use "conscious", "consciousness", etc. to talk about *phenomenal* consciousness. Henceforth I usually avoid the term "self-consciousness", replacing it by "auto-consciousness", because of the non-egological tendency at the base of this article. I want to avoid the notorious ambiguity of "self", as between being an egological noun referring to some sort of holder or agent of experiences but different from them and being a non-egological linguistic prefix tantamount to "itself", as in "self-cleaning oven". But "auto-consciousness" is still meant to include both egological and non-egological possibilities. For the sake of consistency, I will henceforth usually replace the standard term "pre-reflective self-consciousness" (PRSC) by *"pre-reflective auto-consciousness"* (PRAC). I will also refer to reflective auto-consciousness as RAC.

I take a "reflective" mental/brain entity state (state, activity, etc. to be one that deploys concepts, predications, other propositional constructions, and/or reasoning acts. While "pre-reflective" (PR) means not using any such tools, I will be predominantly concerned with the non-using of concepts, so [pre-]reflective will often amount essentially to [non-]conceptual.

I use "reflective" and "pre-reflective" to qualify not just types of auto-consciousness but also types of *externally-directed consciousness*—consciousness of external matters (or matters that were external, could be external, are impossible but proposed as external, or seem external, etc., as in remembering, imagining, hallucinating, dreaming, etc.—I will often refer to the matters it is / seems to be directed at as being "external" , for brevity, even if they are just imagined, say.) However, I do not stipulate that a mental state or process has to be conscious in order to be reflective, so to that it extent it departs form core everyday notion of reflecting upon something, and arguably from the usage of many consciousness researchers.

I avoid assuming that a reflective state/activity involves any strong notion of focusing attention on the target to the exclusion of other matters. So, while "introspection" or self-"observation" might be used to refer to reflective auto-consciousness where one has turned attention to one's consciousness, putting other matters aside, I leave it open that reflective auto-consciousness more generally can exist

as just one amongst many simultaneous activities of equal standing, or as a background conscious activity while one's mind is more occupied (whether reflectively or otherwise) with other matters.

# 2 The Meta-Causal Account in Brief

## 2.1 Overview

One central motivation underlying the account is to *allow* for the possibility that natural or artefactual beings that are below, and possibly way below, the level of human beings have some form of phenomenal consciousness (be it only very crude forms of pleasure and pain, say, or more weakly, comfort and discomfort). I wish to avoid having a theory of consciousness that rules this possibility out of court by virtue of the cognitively advanced nature of mechanisms, processes, etc. or neural circuitry, etc. that it proposes. This motivation is a main reason or viewing consciousness as pre-reflective at heart, with reflectiveness as an optional add-on, so to speak.

One central assumption in the account is that being-conscious is, in the first instance, an occurrent feature of *processes*—genuine, causally proceeding ones, as opposed to "pseudo-processes" (Dowe 2009).[2] It is not in the first instance a feature of a single state existing at some time, or of a belief etc., or of an organism, etc., but can be derivatively ascribed to such things. So I assume that, given a [genuine] process, in other words a chunk of spatially-extended causally-unfolding activity of the world over some time, it is either conscious or not conscious. (The full account allows for degrees of consciousness, but for brevity I suppress this matter here.)

A process need not involve all aspects of physical state in the region of space-time that it occupies, so for example it might only involve electromagnetic aspects of state (plus related causation). Thus, the world outside the process covers both physical state spatiotemporally outside the space-time region occupied by the process and aspects of physical state that are within that region but not included in the process. I place very few restrictions on the shape of the region. As just one example of what I mean, the region of space-time occupied by the neurons and neural fibres in a person brain's neural network as she walks around would form a very complex spatiotemporal region.

I include, as part of the causation "within" a process, the causation coming into it from the world outside (and also causation going out into the world, though it is the incoming causation that is highlighted in this article). It should also be noted that something within the process may be caused by a combination of inner and outer influences.

In principle, a special case of a process is just a constant physical state in a physical region over some time, if the causal influences are such as to maintain this state as it is. However, I do not make use of this possibility, and I generally tacitly assume that there is state change occurring throughout a

[2] A standard example of a pseudo-process is the moving shadow of a person walking. The trajectory of states of the shadow is imposed from outside itself, so to speak, rather than arising through causation between those states. A genuine processes may, however, be helped along, perhaps continuously, by causation from outside, so in fact the matter is not straightforward. For this article, I rest on the intuitive notion that a [genuine] process unfolds naturally by causation, allowing for "input" causation without having exceptional states being imposed on it from outside.

conscious process. But I assume that is in principle possible for a completely (causally-)isolated process to be conscious. In particular, consciousness does not need any "object" (anything it is "about") outside the process it resides in.[3]

I take being-conscious to be a purely objective matter, not as matter of construal or illusion. Note that I do not take it to be an objective matter how the world is divided up into processes: 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.

I cannot go into the reasons for taking a process-based view, but my view as least roughly reflects process philosophy (Seibt 2013), and accords with one strong trend in the area of theories of consciousness, which is that consciousness is a matter fundamentally of activity rather than static state. To take just some diverse examples, it accords on this score with the Integrated Information Theory of consciousness (Oizumi, Albantakis & Tononi 2014) and the centrality of activity in the work of J.G. Fichte (see Fichte 1982) and in the recent work of Strawson (2017). The latter work, on (pre-reflective) "self-intimation," will play an important role later in this article. Strawson affirms the need for something dynamic in the reflexivity of consciousness, not just static reflexivity as in linguistic items that refer to themselves (through devices such as "this very sentence"). Another thought that has some resonance with Strawson's is to do with the potential of computational artefacts being conscious. Consciousness would presumably be a matter of a computer *running a program*, not just sitting there with a program loaded but not running. This is even if all the subtlety of reflexivity, etc., needed for consciousness is implicit in the program itself. This may sound like a trivial and obvious point, but it is a strand within a formative line of thought leading to the meta-causal account.

That line of thought concerns the hypothesis that it does *matter*, and the question of *why* it should matter, as regards the presence of consciousness, that the states of a conscious process are "run through" (proceeded through) and causally related to each other . Suppose copies of them could somehow be arranged to exist all at once over some interval, rather than arising successively. I assume that that would not constitute a conscious system. If one says that the sequence matters because the states have to arise by causation, one then asks, why should that causation itself matter? Why couldn't the same sequence, imposed on some system by force, so to speak, and therefore forming a pseudo-process (a sort of "shadow" of the original process) be conscious? A common particular version of this idea is a replay of a recording of the successive states of a conscious process. (See Kirk 2005 for a similar thought experiment.) [Elsewhere] I give details of such argumentation, supporting the claim that the internal causation within a process matters for its being conscious, so that consciousness is not just a matter of the states the process runs through. However, for the purposes of this article, it is enough to take this claim as a not-very-surprising working premise.

Extension of the line of thinking led to the idea that *a conscious process's own causation matters because it matters to the process itself*, and, more specifically, that an (uninterruptedly) conscious process has to be, at every moment in its time interval, sensitive in some way to its own inner causation as such, at least the causation in very recent process history, occupying a specific "worm" through space-time, leading up to that current moment. Note here that we are going beyond the obvious point that each state in a [genuine] process is affected by ordinary, base-level causation leading up to the state. What we are talking about is the prior *causation as such* somehow affects the

[3] This assumption is not crucial for my account but simplifies some considerations. In practice, no process can be totally isolated.

current state. Also, it is this (possibly brief) causal (sub-)history *as a unit* that is important: the sensitivity is holistic in not just being the sum of sensitivities to individual causings within the history. (A thought experiment and other argumentation presented [elsewhere] supports and fleshes out this claim. It is essentially to the effect that without this sensitivity, it cannot be an objective matter whether a process is conscious or not, given certain other natural assumptions.)

I assume that, even if at each instant there is sensitivity to only a short segment of recent causal history, 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.[4]

However, I repudiate the option that the proposed sensitivity to recent, past causal history could consist of the current state *representing* that (sub-)history (or prior sequence of states). This is because I strongly suspect that representation cannot be made to be all of fully objective, fully pre-reflective and fully local. (See Egan, to appear, for some discussion on this, and Shea 2018 for a detailed overview of representational approaches. I make detailed comments [elsewhere].) In brief: there are well known problems of uncertainty (= need for a theorist's decision) about what is the cause of a particular representation; complex representational structures are in danger of being reflective; and teleological elements in some accounts not only added further non-objectivity but also means it is not a matter local to a particular physical process whether it is conscious or not. One way to capture some of the issues is that it is difficult to see how a lower animal could *represent* a sequence of prior states and their causal relationships (NB even without assuming it has a concept of state or of causation), or a continuous trajectory of states and their causal binding. It is difficult to see how such a representation could avoid being complicatedly propositional in some form or would be something with structural resemblance to the history, bringing in problems of objectively defining what precisely in the world the representation resembles (it might "inadvertently" resemble all sorts of things—and to make matters worse here, we cannot insist that a representation be fully exact, I would claim).[5]

There are philosophically conceivable possibilities other than representation, such as some sort of diachronic non-causal but nevertheless physical grounding (one version might even mean that the current state *contains* the prior history). But I park this as being at least as speculative and mysterious as my own proposal.

My own proposal is that the proposed sensitivity to prior history is *meta-causal.* In my meaning for "meta-causal" (there are others), it is (to oversimplify on some issues) causation where a causal relatum is itself a causation instance, on the intuitive pattern [A-causes-B]-causes-C, or A-causes-[B-causes-C], or [A-causes-B]-causes-[C-causes-D] (or more complicated patterns with additional partial causes or effects are thrown in). If someone said *"John's causing Bill to cry caused Mary to get angry"* (or more idiomatically: *"John's making Bill cry made Mary angry"*) they would ostensibly be describing a meta-causing (meta-causation instance), alluded to by the "caused", whose cause-side is itself a causing, namely John's causing Bill to cry. Of course, the mere existence of such sentences is not evidence that there is meta-causation objectively in the world, rather than projected onto it as matter of construal, any more than ordinary talk involving the term "cause" implies that the physical

[4] Cf. the point in Schechter (2018) that diachronic unity of consciousness might be a matter of such an overlapping succession of intervals of consciousness.

[5] However, see Quilty-Dunn (2020) on perception in children and animals as arguably including discursive structure.

world objectively contains (ordinary, base-level) causation. But with meta-causation in place at least notionally, one can also have meta-meta-causation in principle, and so on up, i.e., where a causing on the cause side or effect side of a meta-causing is itself a *meta*–causing.

While only a tiny minority of mainstream philosophical research on causation even broaches meta-causation (Kovacs 2019), it receives occasional brief mention (e.g., in Ehring 2009, where it is called iterated causation, not to be confused with chained causation, and in Koons 2009, where it is called higher-order causation). It is also implicit in discussions, especially within power theory, whether people can directly perceive causings (Beebee 2009, Groff 2013)—the point being that, if they could, and if perception involves causation as it probably does, then perhaps meta-causation is in play.

To go back to conscious processes, the meta-causal account requires that the current state be meta-causally affected by the prior causal history at least a little way back in time; that is, that (sub-)history, as a unit in its own right, is a partial cause of the nature of the current state, thereby also affecting process state going forward form there, although equally, and in additional, each later state is also influenced by meta-causation from its prior history. Note in particular that the meta-causation within the process itself becomes an aspect of the causal history of later states, leading to a sort of dizzyingly bottomless, circulating meta-ness of the overall causation (rather than a viciously fatal regress upwards—to be discussed below).

Now, although I signalled above that meta-causation as described in English sentences need not actually exist objectively, for our purposes we do want the meta-causation in conscious process to objectively exist in the physical world. So, causation in general must presumably objectively exist (and not just metaphysically, but physically). In fact, the bulk of the present article is insensitive to what particular objective account of causation in general, and meta-causation in particular, is adopted. But it is important to be assured that such a thing is available, given that physical theory can proceed entirely on the basis of equations that do not mention causation, while metaphysical philosophy of causation does not provide a clear, fully objective widely-adopted possibility. I briefly discuss this difficulty further, and motivate my choice for the physical nature of causation, at the end of section 2.2. It is defined at fundamental physical level, not at the level of ordinary physical objects or even in any level favoured in many consciousness theories such as the neural level.

Finally, while consciousness is cast in the account as being a matter of meta-causation, *there may possibly be other meta-causation in the world with no relevance at all to consciousness.* Having raised the possibility of meta-causation, there is no reason to restrict it to lying within consciousness, even though consciousness was the specific reason for raising it, and 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 (and, as should be obvious, there may be meta-causation quite broadly in the universe without consciousness existing broadly). I comment further on this matter in the Conclusion.

In the next subsection I flesh out more precisely what the account claims, to set the stage for the remainder of the article.

### 2.2 Some Specifics of the Meta-Causal Account

### 2.2.1 The Generic PRAIS Necessity Condition

The main initial working assumption on which the account is based is the following necessity condition hold for consciousness, where PRAIS stands for the *pre-reflective auto-individuating auto-sensitivity* mentioned above, and where "auto-individuation" will shortly be explained.

***Generic PRAIS Necessity Condition***

> In order for a process to be uninterruptedly conscious: At every moment in its time-span, however short or long hat time-span is, the process must be physically sensitive to its own existence so far (at least a little way back in time) in a way that is pre-reflective and differentiates (individuates) the process from the world outside itself.

This condition is labelled as "generic" mainly because it does NOT demand that the sensitivity have or not have any specific form, such as meta-causal or representational, *except, crucially, that the sensitivity NOT itself be defined in terms of consciousness (or phenomenality, etc.)*. It also leaves the nature of auto-individuation unspecified. Notice also that it does not directly talk about sensitivity to causation as such, but just to (a recent segment of) the process as a whole.

Crucially also, while the condition is clearly inspired by the notion of pre-reflective auto-consciousness (PRAC) , *it is justified independently of PRAC*. The pre-reflectiveness arises from a desire to allow for the possibility of consciousness in lower lifeforms, and the auto-individuating auto-sensitivity arises from the lines of thinking summarized and argumentation alluded to in section 2.1.

The auto-individuation amounts the process's being sensitive to itself in a way in which it is not sensitive to the outside world (where what this way is constant across all conscious processes). It is an aspect of the idea that the process's own existence as a process, and hence in particular its own causation, matters to the process. It is not that causation extending more broadly outside the process is what matters to the process. The auto-individuation captures, partially at least, the intuition that a conscious episode is somehow sensitive to (i.e., affected by in some way) 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. To use a metaphor of currents in an ocean, a conscious episode is a current that is sensitive to the existence of just this current rather than to a broader segment of the ocean without distinguishing this current in any way.

Because the auto-individuation requirement arises as part of the "mattering" issues underlying the PRAIS condition, it does not rely on assuming the existence of PRAC. Nevertheless, I believe it captures and makes explicit something that has presumably always been intended in discussions of PRAC or indeed of self/auto-consciousness more generally, that the auto-consciousness is indeed consciousness of this particular consciousness. The consciousness in some sense recognizes its own boundaries.

Of course, this is a metaphorical way of putting it, especially in the PR context. We are of course not requiring the consciousness to "recognize" anything at all in any normal sense of tat term, or be sensitive to boundaries as such, least of all have a concept of own-ness or boundaries. Indeed, in the present article it is possible to see auto-individuating auto-sensitivity as a notion of own-ness and to capture all we need as regards own-ness. Thus, own-ness here does not imply the existence of an

owner or subject that owns the processes causation or boundaries or anything else, except for the process as a whole. This affirms an aspect of the non- egological nature of the account.

The PRAIS condition can be seen in one way to be weaker than assuming that all consciousness involves PRAC, as the PRAIS condition does not state that the "auto" aspect is a matter of auto-*consciousness*. But in another way we have added something that goes beyond PRAC, namely the presumptions that being-conscious is a property of *physical* processes and that the auto-sensitivity is physical.

The PRAIS condition as stated is only a *necessity* condition. (I will shortly be stating a sufficiency partner for it, but it is useful to separate them for various reasons.) Thus, it leaves it completely open what additional conditions might be needed for consciousness, whether in terms of extra restrictions on the nature of the auto-sensitivity, or some other condition entirely.

### 2.2.2. Meta-Causal Version of the PRAIS Necessity Condition

Further argumentation, alluded to in section 2.1, leads from the assumption of the Generic PRAIS Necessity Condition to an assumption of the following more specific, meta-causal version.

***Meta-Causal PRAIS Necessity Condition***

> There is a uniform way W of being meta-causally sensitive to [objective, physical] causation such that: For a process to be uninterruptedly conscious: At every moment in its time-span, however short or long that time-span is, the process must be meta-causally sensitive in way W to the causation within itself that occupies some (possibly very short) time-interval abutting the current moment, and must be NOT be meta-causally sensitive in way W to causation that is outside the process but is weakly contemporaneous (intersects the process's time-span).

"Abutting" a time means coming right up to but not including that time.

The auto-sensitivity is now a matter of past causation within the process having a direct causal influence on the current state, where that influence is therefore meta-causal. The auto-individuation lies in the lack of W-type meta-causal sensitivity to contemporaneous causation outside. Notice that outside causation does *not* include causation into and out of the process, as this is covered for convenience by the notion of causation "within" the process, as explained in Section 2.1. (The time-span qualification at the end can be ignored for now. It is included to allow a conscious process to be a temporal segment of another conscious process.) Note that the specification of auto-sensitivity to be a matter of sensitivity causation as opposed to sensitivity to processes has the effect of regarding the identity of a process as defined by its inner causation.

The way W remains to be elucidated in detail, and is included mainly to keep options open. One might consider, more simply, requiring there to be no meta-causal sensitivity of any sort to outside causation . But there seems to be no reason to make such a strong requirement. This is especially so in the light of the comment that there could be non-consciously-related types of meta-causation in the world. A conscious process might thus have meta-causal interaction of those types with causation outside itself.

Finally, I should note that I have subtracted from the necessity condition the particular nature of physical causation that I subscribe to and sketched in Section 2.2.6. As stated it is thus a weaker condition than the one I actually impose elsewhere.

### 2.2.3 The Sufficiency/Constitutivity Conjecture and Type-Identity

Once reasons had been developed for thinking that meta-causal PRAIS is necessary of consciousness, it was then natural to wonder whether it's also sufficient. In fact, I do *conjecture* some suitable form of it to be so, emphasizing the word "conjecture" because I have no arguments for it from premises in the way I have for the necessity condition. However, I believe it is methodologically and intuitively at least a plausible conjecture, for reasons I briefly mention in the Conclusion.

The statement of the conjecture involves the notion of a slimmer process within a given process P. The slimmer process is one that is not P but occupies the same time interval and all of whose physical state is part of the physical state of P (with the implication that the spatio-temporal region of the slimmer process is contained within that of P).

**Sufficiency [in fact Constitutivity] Conjecture**

> Having, throughout its time-span, some suitable form of meta-causal PRAIS (in the way W mentioned above), and not containing a slimmer process that also has such PRAIS throughout, is enough to make a physical process uninterruptedly conscious in at least some pre-reflective, basic phenomenal sense.[6]
>
> Furthermore, this is not about mere logical sufficiency. Rather, the meta-causation within the process that provides this auto-sensitivity IS the process's consciousness. (Or: possessing the auto-sensitivity constitutes the process's being-conscious.)

So the meta-casual account is a type-identity theory of consciousness, in that it literally identifies (uninterrupted) phenomenality (= being-conscious) with possessing meta-causal PRAIS throughout. However, unlike traditional, mind/body type-identity theories, the identity of phenomenality is not specifically with (states or bits of) brains or bodies. The identity is with a type of physical state of affairs that could arise in many different sorts of physical entity, not necessarily confined to lifeforms or anything structurally similar to brains or bodies. So according the account, consciousness is massively multiply-realizable (taking identity as an extreme form of realization).

The "some suitable form of" qualification is included because of course it might be that not just any type-W PRAIS would be enough. As with type W, it is open what the suitability might consist in, other than implying type W.[7]

[6] There is an extra detail that should be included because of relativistic concerns about the speed of causal influence, but these can be ignored for the purposes of this article.
[7] In the mathematical formulation [published elsewhere] there is a suggested technical requirement called "centred reflexivity", which focuses the meta-causation in an intuitively sensible way.

The slimmer-process point is included to prevent it being the case that, for instance, a set of entirely unrelated conscious processes could turn out to be conscious according the account. It also prevents , say, the entire activity process in one's whole body constituting a conscious process just because there is a slimmer process in one's brain that does so.[8]

There is an important point about the phrase "at least some pre-reflective, basic phenomenal sense". This is intended to mean that the consciousness does have a PR aspect, and may be entirely PR, but may *also* have reflective aspects.

I do not 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 (be construable as including) instantiation of concepts, manipulation of propositional structures, and so forth. I further assume that for these reflective aspects to amount to reflective aspects of *consciousness* they must be suitably yoked to the meta-causation that constitutes the particular consciousness at hand. I have barely started work on this topic, but my hunch is that the yoking must be tight and thorough, e.g. going deeply down through a complex propositional structure.

### 2.2.4 Adjuncts to the Sufficiency Conjecture

If one posits the Sufficiency Conjecture is true, there are riders or adjuncts that can be added, giving more specific forms of the meta-causal account. They are not arbitrary, but themselves motivated by the thinking leading to the Meta-Causal PRAIS Necessity Condition. Three such are as follows, building on each other. They are appealed to in some discussions below.

**Adjunct A: Existence of Core Meta-Causation:**

There is a unique, core form of the suitable meta-causal PRAIS that is mentioned in the Sufficiency Conjecture. That is, every conscious process involves this core form.

**Adjunct B: Core Meta-Causation as Core Consciousness:**

The presence, within a conscious process, of the core meta-causation in Adjunct A constitutes the possession by the process of some group of core feelings (which do not involve the use of reflection). Thus, these feelings are present within all consciousness, and are accordingly said to form core consciousness. Core consciousness is thus entirely pre-reflective.[9]

### 2.2.5 Bottomless, Internally Reflexive Meta-Causation

At any point in a conscious process, past causation in an abutting stretch of the process meta-causally affects state at that point. But the same observation applies to all earlier points in the process. So the causation meta-causally affecting state at a point might include that very same sort of meta-causation. In fact, I suggest that it should contain it, for fear of the process not really being sensitive to the crucial meta-causation that makes it a conscious process. So the mentioned meta-causation affecting

[8] This is not the only way of preventing undesirable consequences of these sorts, but is convenient for the present article. I consider doing without it in Section 4.4.
[9] Some specific core feelings are proposed [elsewhere].

state at a given moment is some sort of meta-meta-causation—and of course that very sort of causation is present throughout earlier stretches of the process. So we have some sort of meta-meta-meta-causation, and so on indefinitely. This might suggest a vicious infinite regress of levels.

But I hold that this is not the only view we need take, and in fact is difficult to take, because, assuming that time is continuous, there is no discrete sequence of moments in the process, so there cannot be a discrete "ladder" of levels of meta anyway. Some more advanced and subtle view is needed. I believe this is that there is meta-causation that has no defined height—or rather no defined "depth," as a better metaphor. We are looking down into a bottomless whirlpool, not up a topless ladder. The causation that takes part on the cause or effect side in a given instance of such meta-causation itself has no defined depth. Such meta-causation can just work on such meta-causation without implying a new type of meta-causation.

The reflexivity here is "internal" in the sense that all the meta-causation involved in making the meta-causation reflexive is itself within that very meta-causation—it does not, so to speak, go outside of it and back in. So, as a shorthand I say that consciousness is a matter of ("suitably" restricted) *internally-reflexive meta-causation*, omitting the bottomlessness for brevity.

I believe we just have to accept some such bottomless sort of reflexivity to make sense of consciousness, and it may be part of its apparent conceptual weirdness for us. It is important to note here that, in a mathematical formulation of the account elsewhere, this bottomless reflexivity is precisely codified in the form of system equations that explicitly mention causation of the type I advocate. So what I am proposing is mathematically, and I would claim metaphysically, possible, even if difficult to understand.

The bottomless whirlpool must be present from the beginning of the time-span of the conscious process. However, this does not prevent its being graded in intensity in some way, and for the intensity to increase smoothly from zero in an early part of the time-span.

### 2.2.6 What is Causation in the Account?

It is difficult to find a notion of truly objective, physically real causation in physics, or in the philosophy of causation (for reviews see Ehring 2009, Kutach 2014, Schaffer 2016), especially given that we require the notion to allow the definition of meta-causation. For a start, there are overtly projective accounts of causation, whereby it is merely we who construe events (etc.) as being related by causation. But non-projective accounts rely on such notions as counterfactuals, interventions, classes of events, or mechanisms in which physical quantities such as energy or momentum are transferred. While I cannot argue the case here, such notions all bring in some element of construal or stipulation by people, including a carving up of the world or a conceptualization of it in a certain way, and are not thoroughly objective. And few accounts are open to the addition of meta-causation. The easiest in this regard is where causation is a relation between facts, because then causings are just more facts, and facts can be about causings, but to make this objective and physical would require a view that the world is made up of (construal-free) facts, an idea I find difficult.

As a result I have opted for 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. [10] I give the idea a radical twist, I believe, by making dynamism itself a first-class citizen of the universe (on a par with familiar physical quantities such as mass), with spatiotemporally located "chunks" of it able to interact with other entities in a way that is governed a new, meta-dynamical, type of physical law.[11]

This particular choice for the nature of physical, objective causation is not relied upon in the rest of the present article. There may in principle be other options for a fully physical and objective form of causation and, in particular, meta-causation.

# 3 From PRAIS to (Restricted) PRAC

Here I argue that the meta-causal account of consciousness plausibly implies that any conscious episode possesses pre-reflective auto-consciousness (PRAC) throughout. More precisely, I present arguments for the claims listed below. Bear in mind that, in these claims, for consciousness to be "at least pre-reflective [of X]" means that it *includes* PR consciousness [of X], but may also include reflective consciousness [of X] yoked to it. The possibility or otherwise of this will be further discussed below.

(AC1) Any conscious process is (at least pre-reflectively) conscious, throughout, of at least some aspects of *its own inner causation* (both incoming and entirely inner). (Because of the PR-ness, this does not involve conceiving the causation *as* causation or *as* its own. Analogous lack of conceiving-as apply in the remaining claims.).

(AC2) Furthermore, with an extra assumption about how complete the auto-sensitivity is in conscious processes, we have that any conscious process is, throughout, (at least PRly) conscious of its prior "consciousness-mandated" auto-sensitivity, i.e., *that prior meta-causal auto-sensitivity within itself that is necessary (according to the PRAIS Necessity Condition) for the prior consciousness*.

(AC3) Keeping that extra completeness assumption, we have that any conscious process is, throughout, (at least PRly) conscious of *its own ongoing consciousness so far*. Thus, the process has PRAC. This auto-consciousness cannot be the consciousness of a slimmer process (see Section 2.2.3), so is just an aspect of the overall consciousness.

(AC4) It is unclear whether the PRAC arising in AC3 can be said always to include, in particular, PR consciousness of its own *pre-reflective* consciousness as an item separated out from the overall consciousness. PR consciousness of own *PR* consciousness could be labelled for brevity as *auto-pre-reflective consciousness* (APRC). Nevertheless, it is plausible that there are important circumstances in which the PRAC includes (or just is) APRC. A trivial case is when the consciousness is entirely PR.

AC4brings out that it is beneficial to identify restricted versions of PRAC, such as APRC. The general notion of PRAC hides significant distinctions.

[10] Adoption of the dynamism or oomph view of causation for the purposes of the meta-causal account of consciousness should not be taken as a comment on the worth of other views for other purposes.

[11] [Elsewhere] I lay out the details of 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. In fact I provide system equations for a toy system that contains meta-causation (meta-dynamism) of the sort needed for consciousness according to the account

Claims AC1 to AC3 do *not* depend on the Adjuncts to the Sufficiency Conjecture, but only on that Conjecture together with the meta-causal PRAIS Necessity claim.

**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 causation, by the PRAIS Necessity claim. As mentioned above this "own causation" is intended to include the causation coming in from outside, such as via visual processing of a rose. So, this meta-causal influence of the incoming causation from the rose is part of what makes the process conscious of the rose. The difference from an ordinary, non-conscious process that is causally influenced by the rose is that in this case there is just ordinary, base-level causation affecting the progress of the process.

Now, the simplest assumption to make here is that there is no fundamental difference in the way the causation from the rose meta-causally influences the progress of the process from the way the process's entirely inner causation meta-causally influences the progress. In other words, as regards the meta-causation, it does not "care" whether it is carrying an influence from incoming causation or from internal meta-causation. Indeed, the inner meta-causation is itself directly or indirectly modulated by meta-causation from the incoming causation. Thus, at any moment, the process is being affected both via incoming base-level causation from the rose, via meta-causation from that base-level causation, and via meta-causation from its own inner causation, itself modulated by the input from the rose. I submit that it would be *unmotivated* toad the *extra complication* of suppose that the fact that the process is conscious *of the rose* as opposed to something is only a matter of the first two types of influence, and not the third. All three are all intimately combined. There is no reason to complicate the picture further by assuming that the process is differentially affected, through meta-causation, by its own entirely inner causation as opposed to other causation. If this is right, then there is just as much reason to say that the process is conscious of its own inner causation as of the rose.

A slightly different way to put the argument is as follows. The progress of the process has, on the "cause sides" of its meta-causings, both the incoming causation from the external matters and its own meta-causation. The incoming causation and the internal meta-causation are co-causes, via meta-causation, of the precise way the process proceeds. That meta-causation acting as a cause is itself, of course, the precise way it has turned out to be partly because of the incoming causation. The process is meta-causally affected by the incoming causation partly by virtue of being meta-causally affected by its own meta-causation considered as cause.

Note here that the Sufficiency Conjecture only mentions sufficiency for consciousness in "at least some pre-reflective, basic phenomenal sense". So there's no immediate imperative to conclude that the process's consciousness of its own causation is reflective. However, since the consciousness of, say, a rose may often be reflective, the comments above, relying as they do on parity between the rsoe and the inner causation, naturally raises the question of whether the consciousness-of-own-causation might often be reflective. This is improbable, of course, and I discuss the matter briefly below.

In summary, we have argued that considerations of *simplicity* lead plausibly to AC1. It would be additional complication, beyond the necessary complexity inherent in PRAIS, to ensure that the conscious episode was *not* conscious of its own inner causation. Thus, I go in precisely the opposite

direction to any idea that auto-consciousness is an extra complication. It is also useful to summarize the argument as making the following points, which are duals of each other:

(i) being conscious of an outer X is just an aspect of the process's particular *auto*-consciousness on the current occasion, modulated as it is by input from X; and

(ii) being auto-conscious in the particular way it is on a given occasion *is the process's way* of being conscious also of various outer Xs.

Consciousness of outer X and the auto-consciousness involved are two sides of the same (meta-)causal coin. As a straightforward special case, if there is no outer X that the process is conscious of, it is merely auto-conscious.

**The Argument for AC2**

The extra assumption, about how complete the auto-sensitivity is, mentioned above for AC2 is that: the process's consciousness-mandated direct, meta-causal sensitivity at time *t* to its prior causation includes direct sensitivity to, in particular, the prior meta-causation constituting the prior consciousness-mandated PRAIS.

I now appeal to a parity argument with consciousness of an external rose.

Suppose one is looking down on the rose, and can therefor see a middle part of the overall rose-part that one is seeing. That is, the middle is a partial cause of one's conscious visual processing of the rose. I claim that this is enough to say that one is at least PR conscious of the middle part. One may not separately conceptualize the middle part even if one is reflectively conscious of the rose, but that does not matter for present purposes. All that matters is that one's conscious state is *as directly affected by* the middle part as it is by other parts. This example involves a spatial division of the middle part from the rest, but we can easily vary the example to be about, say, a fine-scale colour pattern that suffuses the whole visible part of the rose, and which as directly affects one's particular state of consciousness as other aspects of the rose do. We can then say that the person is at least PR conscious of the patterning.

Equally, given that the above-mentioned meta-causal aspect of prior inner causation just as directly affects the particular consciousness going forward as any other aspect of that causation does, we can say the process is at least PR conscious of that meta-causal aspect.

In sum, the structure of the argument is that the prior meta-causation, while merely an aspect of the overall prior causation, has just as direct an effect on ongoing consciousness as other inner causation does and as the rose does, and hence it can be viewed as something the process is at least PRly conscious of, given that it can be regarded at least PRly conscious of the inner causation more generally and of the rose.

**The Argument for AC3**

Here I appeal to the following principle, which emphasizes that, while PR consciousness is a mental matter, it is not complicated by matters such as conceptualization and modes of presentation:

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

From AC2, we have that the conscious process is at least PRly conscious of the above-mentioned meta-causal aspect of prior causation. It possesses PR consciousness of that meta-causation. But by the identity thesis inherent in the Sufficiency Conjecture, that meta-causation just is the prior consciousness possessed by the process. So, by the De-Re Principle, the processes possesses (at least) PR consciousness of the prior consciousness right up to the current moment. (Consciousness is not defined in the present account at instants, so the omission of the current moment does not miss out any distinctive instance of consciousness.) The process possesses PRAC.

**Discussion of AC4**

Can we infer from a conscious process being (at least PRly) conscious of its prior consciousness (in all its aspects, a little way back in time), which is what the AC3 argument delivers, that it is (at least PRly) conscious of its prior *pre-reflective* consciousness (a little way back in time). In other words, can we infer that it possesses some *auto*-pre-reflective consciousness (APRC),as a special case of PRAC?

After all, it may be that there is an argument available like the one used for AC2. Let us suppose that:

(PRIDENT) the prior PR consciousness can be identified with a special "PR aspect of" the AC2 meta-causation (the prior meta-causation that the AC2 argument is about: the consciousness-mandated prior meta-causation to which the process is required to be sensitive by the PRAIS Necessity claim).

That is, we are supposing that that meta-causation has a special, objectively identifiable aspect that constitutes a PR aspect of the prior consciousness. This aspect might encompass all of the prior PR consciousness.

Given this supposition, we might argue that a PR aspect of the AC2 meta-causation has just as direct an effect on the consciousness going forward as other aspects of the AC2 meta-causation do. Then, by parity with the way we argued in AC2, we should be happy to conclude that the process is (at least PRly) conscious of its prior PR meta-causation aspect, and hence (by identity as in the AC3 argument) with (at least PRly) conscious of some or all of its prior PR consciousness.

However, I have no particular, strong basis at present for claiming PRIDENT. One consideration that may help with future arguments for PRIDENT arises if we can take Adjuncts 1A and 1B to the Sufficiency Conjecture to be true. Then, the prior consciousness above is somehow or other built upon the core, PR consciousness mentioned in Adjunct 1B, where this core consciousness is the core form of meta-causation mentioned in Adjunct 1A. Depending on what this "built upon" amounts to, core meta-causation may be a ready-made or consequent candidate for a "PR aspect" in PRIDENT.

But, even so, the prior PR consciousness may well include more than the core consciousness. So it would only have been established that the process is PR conscious of *the core aspect of* its prior PR consciousness. Still, this may be enough for us to say that it is PR conscious of its PR consciousness (i.e., it has APRC). It seems compatible with many prior discussions of forms of auto-consciousness that it is not necessarily consciousness of the consciousness in all its aspects.

**Additional Observations**

If we had not taken the auto-sensitivity in the PRAIS that is necessary for consciousness to consist of *meta-causation*, then we would not have had a path of argumentation to AC3. The argumentation rests crucially on the fact that both the rose and the process itself directly affect the process causally (where the self-affect is meta-causal because of that directness). If, for instance, consciousness were claimed to be a matter of reflexive *non*-causal auto-sensitivity, we would not have been able to claim that the process is conscious of its own auto-sensitivity by virtue of parity with being conscious of the rose.

I am sceptical about intentionality (about-ness) in general being fully naturalizable (cf. in particular the comments on representation in Section 2.1), and therefore take the "of" in "conscious of X" where X refers to something outside the consciousness generally to be partly a matter of pragmatic construal by us as ordinary people or philosophical commentators. This stance is because I assume that causation from X is importantly involved somehow, and there are familiar problems such as why we say it is X in particular that has exerted the causation on the conscious process as opposed to something closer or further out on a relevant causal chain. However, in case of a process being conscious of something internal, be its own causation or consciousness, the meta-causal account gives us the ability to take the "of" fully objectively. This is because we do not need to stop on causal chains within the process: anything on a chain that is involved in the process's consciousness can be viewed as part of what the process is (at least PRly) conscious.

In arguing for AC1—3 we left open the possibility that the auto-consciousness being argued for might in principle have reflective aspects as well PR aspects—in other words, might in principle include RAC (reflective auto-consciousness), and perhaps even reflective consciousness of own causation or particular meta-causal aspects (in for instance having concepts of those aspects). I do leave this as an in principle possibility, and perhaps consciousness in advanced non-humans (artefacts, aliens, …) could be full of such reflective consciousness including RAC. But as a pragmatic, contingent matter, I assume that human and less advanced beings do not naturally have the conceptual ability to have reflective consciousness of the causal matters discussed above, and even if equipped in principle to have RAC, RAC would normally be extra, unproductive activity when the being is engaged in activity in the everyday world.

# 4 Other Consequences and Ramifications of the Meta-Causal Account

## 4.1 Intrinsic Self-Intimation

This article's approach has a strong affinity with Galen Strawson's theorizing about self-intimation (2017). Self-intimation (which I will equate to PRAC in this article) is intrinsic to experience, rather than being an *extra* awareness (separate component of one's conscious state). (See also Textor 2015, Montague 2017.) Consciousness "comports" consciousness of itself, where "comport" means 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."[12]

Similar claims emanate from the theory in this article. Crucially, having phenomenality at all is equated with having a suitable pattern of internally reflexive meta-causation, in the sense explained in Section 2.2.5. This pattern can be modulated, 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 reflexive meta-causation. But having this pattern is, just as much, being conscious of this very meta-causation, by the arguments in Section 3. So the process is intrinsically and inseparably both conscious of the external object and 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, from this article's approach we see that there is a sense in which the reflexivity of consciousness is what actualizes consciousness from moment to moment. But with our meta-causal unpacking of consciousness, we can get away from the mysterious-looking circularity of this statement. Rather, it is the reflexivity of the proposed type of *meta-causation* that actualizes consciousness from moment to moment—where that statement is now non-circular—because consciousness just is a matter of the reflexive meta-causation, which is constantly causing itself to continue.[13] But that reflexivity is foundational to the consciousness. We can take Strawson's somewhat mysterious "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 very (time-extended) reflexivity it sustains itself through time. This article's proposal is very much about PRAC being the "dynamic essence" of consciousness and being an "infinite whirl" of meta-causation.

Because the consciousness of a process is intrinsically and unitedly both of the external matters and of itself, the consciousness of own consciousness that Section 3 argues for means that the second occurrence of consciousness in "consciousness of own consciousness" might be argued to refer in part to the process's own auto-consciousness (though this notion of referring-in-part needs close attention). Thus it might be argued that there is consciousness of consciousness of consciousness. We would then be able to iterate this indefinitely. This is, however, not a vicious infinite regress of layers of distinct states/acts of consciousness, but just a way of having an infinite ladder of alternative ways of describing the same intrinsically reflexive thing.

I also claimed in Section 2.2.5 that the internally reflexive meta-causation constituting consciousness does not involve a metaphysically vicious regress of layers of meta-causation. We just have an infinite descriptive ladder, useful from some heuristic purposes, whose rungs are terms like meta-meta-meta-causation, although we affirmed that the reality was more like a bottomless whirlpool with no distinct layering. However, even if we consider the ladder of meta, it is *not* that it aligns with the above consciousness ladder in the obvious way. Concentrating for simplicity on just PR consciousness, the alignment would be as follows: (1) (PR) consciousness of external X consists of meta-causation M1 with causation-from-X as (partial) cause; (2) consciousness of that consciousness is a meta-causation instance M2 with M1 as (partial) cause; (3) consciousness of that consciousness (the consciousness at the start of (2)) is a meta-causation instance M3 with M2 as (partial) cause; and so forth. Rather, any

[12] In summarizing some of Strawson's claims, I have translated "awareness" as "consciousness" for consistency with remainder of the present article.
[13] An important self-sustaining quality of the meta-causation constituting consciousness is addressed in detail [elsewhere].

level of consciousness of anything involves the *whole* of the meta-causal ladder, describing the overall, integrated internally reflexive meta-causation, rather than just one rung of it, and concomitantly (PR) consciousness of own (PR) consciousness is consciousness of that overall meta-causation, rather than of meta-causation at some particular notional level.

Now, it is true that consciousness from some moment *t* onwards in a process arises because there is some suitable meta-causal "link" lying across *t* that takes the prior consciousness-mandated meta-causation as (partial) cause and influences the process state going forward. And concomitantly the prior consciousness is at least a partial cause of the consciousness going forward. But there is no sense in which consciousness at *t* consists of that *t*-crossing meta-causation. For a start, consciousness is not defined at instants. So, suppose instead we consider a small non-empty interval around *t* and consider the consciousness within this interval. This consists of the whole bottomless "whirl" of internally reflexive meta-causation within that interval and in to which the particular meta-causal "link" across *t* gets tangled up (or absorbed, to use a possibly better metaphor). And there is no "part" of the consciousness that can be identified with that link. In other words, the reflexivity of consciousness arises from the internal reflexivity of the meta-causation, but the relationship between the reflexivities is complex.

### 4.2 For-Me-Ness in the Account

The meta-casual account enables a straightforward explication of the for-me-ness of consciousness, as follows.

In 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 involves phenomenology of for-herself-ness (or perhaps even belonging-to-herself-ness, but I will leave this version to another occasion) This is not separate phenomenology, in that it is integrated with the redness etc.: there is phenomenology of redness-for-herself-ness (This would usually be talked of as for-me-ness and mine-ness, and see notably Guillot 2017 and Zahavi 2018 on the range of different notions available in the vicinity of for-me-ness and mine-ness.)

My claim 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. (Caution: I do not mean "affecting" in the sense of giving rise to affect (emotion, evaluation, etc.), but merely a sense of leading-to-some-effect-in, although this *could* be an effective effect). So her rose-based phenomenology contains phenomenology of redness-affecting-herself, roundness-affecting-herself, silky-texture-affecting-herself, etc., and there may also be phenomenology of relatedness (for herself) of different aspects of the rose with each other. 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 the meta-causal account, 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, 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 external X and consciousness of own consciousness on that occasion are two sides of the same causal and meta-causal coin (see Section 3). The meta-causal account provides an account of qualities such as for-herself-ness that does not rest circularly on talk of consciousness.

In saying this I am downplaying the more "personal" aspects of "me" or "herself", and concentrating on the for-herself-ness being at core a non-egological 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 that is needed for personal herself-ness here may be similar in quality to the envisaged yoking of reflection to PR consciousness, and indeed reflectiveness and that extra for for-herself-ness may overlap, but this issue is beyond the scope of this article.

### 4.3 Transparency Issues

As noted in Section 4.1, the meta-causal account supports the "intrinsicality" thesis that consciousness of something X intrinsically includes at least PR consciousness of (aspects of) that consciousness. Indeed, the core of the argument for this was that, given the structure of (meta-causation) involved in the episode, it is difficult to draw a line between the other-directed and inner-directed aspects of the consciousness. As part of this, they 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 a more direct and local causality is involved. Then there is tension is with the idea that the experience, where for instance Sally is consciously seeing an external rose, is one where there is phenomenology of a *rose* having certain qualities. 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, I think with some intuitive correctness, claim that the phenomenology is more strongly rose-focused than focused on the person or her consciousness. But further (brief) comments are possible here.

First, while there is some sort of parity between the way the incoming causation from the rose and the way the internal meta-causation work, it remains the case that it's the incoming causation that is at the root of the current particular, "rose-modulated" internal internally-reflexive meta-causation. There remains an inherent asymmetry in their involvement, and this could be enough to explain the "rose bias" in the phenomenology.

Secondly, the impression of rose bias arises, I believe, in our consciously reflecting upon our phenomenology. I doubt that we have PR consciousness of rose-bias as such. But to the extent that we reflectively, consciously thinking about our phenomenology, the only conceptual access we may have to anything close to our the involvement of ourselves in the intrinsic make-up of the phenomenology may be precisely via broad concepts of phenomenality or consciousness. But in entertaining that concept we don't conceptualize our involvement within its make-up (as opposed to our mere

possession of the particular phenomenality at hand). That is, in reflecting upon our particular phenomenality on a given occasion, we are unwittingly conceiving of something that lacks the sort of bias that we reflectively consider it to have.

Strawson (2017) touches on what is effectively a transparency issue when he 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" of conscious episodes and the actual, metaphysical "loopiness" of consciousness inherent in its involving self-intimation. 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 the somewhat mysterious notion that the phenomenal simplicity consists in the metaphysical complexity to be clarified. There is the loopy complexity, i.e. the bottomless, internal reflexivity, of the consciousness-mandated meta-causation of a given conscious episode. The episode's phenomenality consists in the episode having this loopily-complex meta-causation. But, as at least suggested by the point in Section 4.1that the conscious whirl has a holistic relationship to the bottomless meta-causal whirl, as opposed to a level-by-level relationship, there is no reason to presume that the loopiness of the meta-causation shows up as *loopiness* of the phenomenality. Rather, we can suggest that it shows up simply as phenomenality itself, where, as an intrinsic aspect of phenomenality, the different types of phenomenal character within the experience are intrinsically for-that-experience (or for-me, if it is legitimate to introduce a "me") or felt by that experience to be its own (or experienced as mine).

### 4.4 Issues concerning the Possible Unity of Consciousness

The unity of consciousness is a large subject (Masrour 2020, Schechter 2018). The main driving force in the issue of unity is that consciousness (allegedly at least) *feels* unified at any given moment: to put it less vividly, this feeling is part of the phenomenology of consciousness. The driving force is not the unity of, say, whatever naturalistic mechanisms might underlie consciousness. But of course the question of how the felt unity arises from underlying mechanisms (or other features of a given account of consciousness, of any sort, physicalist, dualist, …) is also a crucial one. Also, there is both a synchronic and a diachronic dimension to felt unity. This distinction is made murky in a process based view of consciousness where consciousness is defined only over intervals of time anyway. Nevertheless, I will assume that there is a sense both of unified-now and unified-with-at-least-recent-past (where "now" does not presume an extension-less instant of time). For brevity I will address only the former, while recognizing that a full account may not be fully extricable from accounting for the latter.

Approximately following Brentano, Textor (2015) proposes a view of the (synchronic) unity of a person's consciousness at a given time. The view has it that the unity is not a matter of apparent unity of the outer objects to which the consciousness is directed but rather of the mental activity going on at a time—this activity composed possibly of many subsidiary mental activities—with one inner awareness , which is the inner awareness of the whole activity being directed upon that activity. The subsidiary activities do not have their own inner awarenesses directed upon themselves. The uniqueness of the inner awareness, combined with its single-directness, accounts for the feeling of unity.

In one way, the meta-causal account as so far discussed fits with that view, because of the slimness-minimality in the Sufficiency Conjecture (tantamount to saying that a conscious process contains no slimmer conscious process). Thus, in particular, the auto-consciousness of a conscious process could not be the consciousness possessed by a slimmer conscious process. In any case, we have argued that the auto-consciousness is just an inextricable aspect of the overall consciousness. It is natural then to say that there is just one auto-consciousness, even if the constitutive meta-causation can be roughly divided up in some way, for instance modulated by different outside-world matters the process is conscious of.

However, while Textor's analyses are congenial to the meta-causal account in various ways, I am sceptical of the above account of unity. Just saying the inner awareness has a single object does not of itself preclude that object as being complex and of the inner awareness being aware separately of the parts as a *part of* being aware of that single whole. There could be just as much unity or disunity here as there is in how the external matters appear to the consciousness. Of course, one could just stipulate that the inner awareness is more unified than the outer objects are, but it would be better to have an independent justification.

That point aside, the meta-causal account in any case goes in the other direction. It has it that the particular auto-consciousness (inner awareness) on a given occasion is modulated by the effects of the causation coming in objects that the conscious process is conscious of, and that the latter consciousness-of is just the particular way in which the current process is auto-conscious. This matter of being two sides of the same coin encourages the thought that the two sides are as unified or non-unified within themselves as the whole coin is. The degree of felt unity of the outer objects is at least roughly similar to the felt unity of the consciousness as the object of auto-consciousness—if indeed these feelings can be distinguished at all. To put it more directly, because of the modulation of the inner meta-causation by the incoming causation from outer objects, the appearance of unity of the outer will be similar to (and perhaps not divisible from) the feel of unity of the inner. The account as it stands does not, however, predict how strong the feel(s) might be.

There is a rather vacuous way in which the meta-causal account accords with the Textor view, in that a conscious process cannot (according to the Sufficiency Conjecture as stated in Section 2.3) contain another, slimmer one. There is no room in the first place for having, say, multiple inner strands of consciousness (slimmer conscious processes) that separately are conscious of the whole consciousness. However, one might entertain removing that slimness-minimality part of the Conjecture, and adopting some other way of avoiding unwelcome consequences that part presents (such as a collection of independent conscious processes being *ipso facto* a conscious process). One such way might be require the process to have a certain level of internal causal integration, according to an integration measure roughly on the lines of the Phi value proposed by IIT (the Integrated Information Theory of consciousness, Oizumi *et al,* 2014).[14] Thus, for instance, it might be possible to have a conscious process that consists of some slimmer processes that are individually conscious together with an interconnecting web of causation, or a conscious process that consists of a conscious process "surrounded" by additional causal mechanism, where in either case the causation overall is sufficiently integrated. My observation here would be that the more that the interconnecting/additional

[14] However, the measure will not do as it stands, as it has been shown to be non-objective (i.e., partly a matter of human decisions about system structure, etc.). See e.g. Barrett & Mediano (2019). It also does not typically address causation at the microphysical level envisaged in the meta-causal account. A possibility that fixes both issues are quantum-theoretic measures of integrated information, as in Kremnizer. & Ranchin (2015), for example. But these approaches, along with IIT, do not address meta-causation, so would need to be varied to encompass it.

causation is "miscellaneous" in the sense of not being of the sort that is needed for consciousness (perhaps it is merely base-level causation, or perhaps it is meta-causation of a type that is irrelevant to consciousness), the more that the overall consciousness feels disunited, I would claim. Going in the other direction, the more non-miscellaneous it is, the more danger there is that an individual process in the mix will not satisfy auto-individuation (it will be in danger of having W-type meta-causal sensitivity to contemporaneous matters outside itself) and so not be conscious after all. But there may be a compromise that would allow a reasonably well unified multiple-consciousness possibility. (This would correspond to the "many-in-one" possibility of consciousness, Schechter 2018).[15]

It seems very likely, of course, that if a cognitively mainstream, intact-brain adult human can have multiple consciousnesses, only one of them would be egological in any strong sense, otherwise it would probably be common for people to make reports or take actions that would betray that fragmentation. Also, the long-term self of the person would involve the stitching together of the series of egological consciousnesses that the person goes through, with memory crossing any time gaps in between (recall that in this article an individual consciousness is temporally uninterrupted).

Aside from multi-process possibilities, something the account allows is for different strands of processing within the process to be reacting in different ways to the process's consciousness, or possibly reflectively thinking about the consciousness in different ways if there is a reflection capability. I do not know how this would affect the feel of unity, but the point being made is that even PR auto-consciousness might be complex, over and above any complexity encouraged by the process being conscious of a complex external environment. There is no need to think that PR consciousness is a simple matter just because it is internally unified (if it is) and lacks the particular complications of reflection.

# 5 Conclusion and Ongoing/Future Work

I hope to have presented some grounds for supposing that, if the meta-causal account of consciousness or some future variant of is at all reasonable, it provides a fruitful and solid basis for further justification and exploration of pre-reflective self-consciousness (or pre-reflective auto-consciousness as I prefer to call it), including surrounding issues such as: the intrinsicality of such auto-consciousness as an aspect of consciousness; for-me-ness; transparency intuitions; and the unity of consciousness. Indeed, the comments made on these issues all rest on much the same (meta-)causal considerations. The account is expressed in terms of naturalistic matters that do not themselves make any reference to consciousness. In particular, we have seen how one can plausibly infer that consciousness includes pre-reflective auto-consciousness from an assumption that it includes a type of reflexive auto-sensitivity that is not itself defined in terms of consciousness. The hope is that these developments will not only advance the philosophy surrounding pre-reflective auto-consciousness but also turn around to act as evidence that the meta-causal account is on the right track.

Indeed, it was the assumption of reflexive auto-sensitivity together with the assumption that consciousness is a fully objective aspect of the physical world, and one that does not intrinsically require reflection, that led me to the idea of meta-causation as an objectively existing aspect of the

[15] If consciousness and W-type meta-causation is graded in intensity, in a systematically related way, then it is possible to have a conscious process that is relatively weak overall containing stronger conscioius processes. I omit the description of this for brevity.

physical world. This then required an objective, physical form of causation in general, and prompted the adoption of a basic-level physical dynamism in the world as that form. All in all, a consideration of the intimate reflexivity of consciousness led to a suggested radical new view concerning causation in the physical world on the one hand and a new detailed mechanistic framework in which to couch consciousness on the other.

On the former, physical side, recall, however, that the arguments in section 3 onwards do not rely on the particular equation of causation with the dynamism or "oomph" of section 2.2.6. In principle, there may be some other choice for the nature of causation that would serve the arguments equally well, but it does have to be fully physical and objective and furthermore to allow the fully objective definition of meta-causation.

Meta-causation is a radical thing to adopt in metaphysics, though not unprecedented, and even more radical and possibly completely unprecedented as an addition to physics. It is also highly weird intuitively, especially in the particular form of a "bottomless whirlpool" of internally reflexive meta-causation. I see this as a feature, not a bug, because of the weirdness of consciousness itself, when viewed alongside the physical world. I am not just arbitrarily conflating two weirdnesses in the world to reduce their number. Rather, I believe there is a principled similarity between the two weirdnesses that deserves further exploration. To put it briefly and provisionally, I submit that the internally reflexive meta-causation has just the sort of self-interiority that we sense consciousness as having and as making it so mind-bending. The self-interiority of internally reflexive meta-causation is much tighter than what one gets from trying to make a complex, articulated system include a feature that counts as self-representation (even putting aside worries about whether representation has any fully naturalistic nature). Rather, meta-causation provides us with a sort of self-interiority that is tightly wound on itself in a deeper sense. Or perhaps a better way to put it is that the self-interiority is all the way down, whereas self-representation in an articulated system is a whole-system property, not something one sees as one probes the innards of the system. Also, I am bound to say that, while one can certainly challenge the meta-causal account with the question: why on earthy should that meta-causation constitute phenomenal consciousness? One can ask the analogous question even more fiercely and, I would say, with less hope of a satisfying answer, of any other existing physicalist theory: to take just one example of many, why on earth should making things interact in a global workspace (Baars 1988) create phenomenality?

These remarks are related to profound connections that I see between the meta-causal account with ideas from certain past researchers, notably Fichte (1982). I have only just started to look at these connections, but the core of the matter in the Fichte case is his idea that action that intrinsically acts upon itself is the primitive basis for everything including consciousness (though some features of his ideas on consciousness conflict with the meta-causal account).[16] As one roughly, possibly Fichtean thought, I am tempted to propose that a certain extreme version of the bottomless whirlpool of meta-causation is able to exist, at least in principle: *viz*, an isolated pure consciousness consisting just of core consciousness, with no base-level causation at all. There is just (suitably restricted) meta-causation acting upon itself, and indeed where the meta-causal self-acting is itself.

[16] Preliminary connections are expressed [elsewhere]. I am tentative in drawing connections to Fichte given that his work is notoriously difficult (Breazeale 2018, Henrich 1982, Zöller, 2016).

There are many other issues for ongoing and future research based on the meta-causal account. Some of these implicitly or explicitly involve pre-reflectiveness, such as the issue of how consciousness can arise in biological evolution and in the development of an individual. The account also suggests new lines of thought on time consciousness and the metaphysics of time.

## References


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

Barrett, A. & Mediano, P.A.M. (2019). The Phi measure of integrated information is not well-defined for general physical systems. *J. Consciousness Studies, 26* (1--2), pp.11--20.

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

Breazeale, D. (2018). Johann Gottlieb Fichte. In E.N.Zalta *et al.* (Eds), *Stanford Encyclopedia of Philosophy*, Summer 2018 ed.

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

Dowe, P. (2009). Causal process theories. In H. Beebee, C. Hitchcock & P. Menzies (Eds), *The Oxford Handbook of Causation,* pp.213--233. Oxford University Press.

Egan, F. (to appear). A deflationary account of mental representation. In J. Smortchkova, K. Dolega & T. Schlicht (Eds), *What are Mental Representations?* Oxford University Press.

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

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

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

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

Henrich, D. (1982). Fichte's original insight. (Trans.~David R. Lachterman.) In D.E. Christensen (Ed.), *Contemporary German Philosophy: Vol. 1*, pp.15--53. University Park, PA: Pennsylvania State University Press.

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

Koons, R.C. (1998). Teleology as higher-order causation: A situation-theoretic account. *Minds and Machines, 8***:** 559–585.

Kovacs, D.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.

Kremnizer, K. & Ranchin, A. (2015). Integrated information-induced quantum collapse. *Foundations of Physics, 45*, pp.889--899.

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

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

Masrour, F. (2020). The phenomenal unity of consciousness. In U. Kriegel (Ed.), *The Oxford Handbook of the Unity of Consciousness*,pp.208—229. Oxford University Press.

Montague, M. (2017). What kind of awareness is awareness of awareness? *Grazer Philosophische Studien, 94*, pp.359--380.

Oizumi, M., Albantakis, L. & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Computational Biolog, 10* (5): e1003588.

Quilty-Dunn, J. (2020). Concepts and predication from perception to cognition. *Philosophical Issues, 30*, pp.273--292.

Schaffer, J. (2016). The metaphysics of causation. . In E.N.Zalta, (Ed.),*The Stanford Encyclopedia of Philosophy*, Fall 2016 edition.

Schechter, E. (2018). The unity of consciousness. In R.J. Gennaro (Ed.), *The Routledge Handbook of Consciousness*, pp.366--378. Routledge.

Seibt, J. (2013) Process philosophy. In Zalta, E.N. (Ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2013 ed.).

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

Strawson, G. (2017). Self-intimation. In G. Strawson, *The Subject of Experience*, chapter 8. Oxford Scholarship Online, March 2017.

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

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

Zöller, G. (2016). Fichte's original insight: Dieter Henrich's pioneering piece half a century later. In K. Gjesdal (Ed.), *Debates in Nineteenth-Century European Philosophy: Essential Readings and Contemporary Responses*, pp.45--56. Routledge.