

PHENOMENAL TIME AND THE METAPHYSICS OF THE MIND

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Lyu Zhou

August 2023

© 2023 Lyu Zhou

PHENOMENAL TIME AND THE METAPHYSICS OF THE MIND

Lyu Zhou, Ph.D.

Cornell University 2023

My dissertation concerns phenomenal time, i.e. how time appears to us subjectively. The key theme of my dissertation is that thinking about how time appears to us subjectively helps us answer many classic metaphysical questions about the nature of time and human consciousness.

Chapter I argues that our mind imposes time upon our consciousness as its essential constitutive structure: in Immanuel Kant's (1781/1787/1996) language, phenomenal time is an *a priori* form of our consciousness. I argue that our consciousness is necessarily temporal, and this necessity has an important revelation: our mind does not perceive time, because perception is a causal process, and yet no causal mechanism, due to the contingency of its operation, can ensure that our consciousness is necessarily temporal. Instead I propose that our mind imposes time upon our consciousness as its essential constitutive structure. My proposal leaves open the question of whether the world as it is independent of our conscious experience is temporal at all.

Chapter II argues that given how time appears to us subjectively, our consciousness cannot be purely physical. Our immediate present consciousness – what William James (1890/1950) calls the *specious present* – has a (non-instantaneous) duration. I argue that this specious present is a phenomenally extended unit of consciousness that is mereologically inverted in the sense that the parts depend on the whole: the shorter constituent durations of the specious present cannot

exist except as parts of the whole specious present. Yet what is physical – a physical object, process, or functional system – does not have this peculiar property of mereological inversion: instead, any physical whole depends on its parts. Therefore, given such a structural discrepancy, our specious present cannot be identical to, or purely constituted by, what is physical.

Chapter III defends the methodology of conscientious introspection employed in the preceding chapters. After clarifying Uriah Kriegel's (2015) helpful distinction between the reliability and the potency of introspection, I argue that a full appreciation of this distinction has important revelations: one is that many of the pessimistic concerns with introspection threaten not so much the reliability as the potency of introspection; and another is that, once the two notions are disentangled, the reliability of conscientious introspection emerges as eminently defensible.

BIOGRAPHICAL SKETCH

Lyu Zhou does not know much about anything. After completing his B.A. *summa cum laude* in Philosophy and Classics at Columbia University in the City of New York in 2016, he joined the Sage School of Philosophy at Cornell University in the same year to try to know a bit more with the help of the sages. There he was pleased to be awarded an M.A. in Philosophy in 2019 after a grilling oral exam. Now he is studying phenomenal time and consciousness, with the hope that he might become less ignorant about (a) the nature of human mind and personal identity, (b) the possibility of human free will and moral responsibility, and (c) the nature of divine consciousness and its relation to time. He is an avid reader and admirer of the history of philosophy: Immanuel Kant, Augustine of Hippo, Plato, and Simone Weil are his intellectual heroes. He is very grateful that his advisor, teachers, and colleagues are so extremely kind and patient with his silly annoying questions.

To my parents, with love and gratitude.

Δόξα ἐν ὑψίστοις Θεῷ
καὶ ἐπὶ γῆς εἰρήνη
ἐν ἀνθρώποις εὐδοκίας.

ACKNOWLEDGEMENTS

I would like to express my enormous gratitude to my advisor, Derk Pereboom, an exemplary intellectual and moral paradigm in every respect. His encyclopedic knowledge and profound insights are and will remain a lifelong inspiration for me. Without his kind help and generous encouragement, I could not even imagine finishing this dissertation. I would also like to thank the other members of my special committee, Nico Silins, Carlotta Pavese, and W. Starr, for their bounteous support. Their keen and constructive feedback truly helps me grow as a thinker. My greatest debt is to my parents, to whom I dedicate my dissertation. Their unfailing love is what makes me who I am. I am also grateful for Lin Zhang's friendship. I fondly remember our philosophical and theological conversations.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH

v

DEDICATION

vi

ACKNOWLEDGEMENTS

vii

CHAPTER I

Phenomenal Time and the Constitutive Structure of Consciousness:

A Constitution Account for Phenomenal Temporality

1

CHAPTER II

Specious Present, Phenomenal Extension, and Mereological Inversion:

A Problem for Physicalism about the Mind

36

CHAPTER III

Introspective Pessimism and the Reliability and Potency of Introspection:

Towards a Plausible Principle of Reliable Introspection

72

CHAPTER I

Phenomenal Time and the Constitutive Structure of Consciousness:

A Constitution Account for Phenomenal Temporality

I. Introduction and Map of the Chapter

On my computer runs Anton Bruckner's 8th Symphony. The brass instruments are sounding triumphant notes in rapid succession, building into a glorious coda. Meanwhile, as the first sunlight sifts through the drawn shades, I am welling up with joy and warmth, with my old anxiety quickly fading away. As this example illustrates, our phenomenal consciousness has an important feature: temporality. Everything of which we are conscious appears in time. I will use 'phenomenal time' to denote time as it appears to us in our consciousness. All our *conscious contents* – namely, whatever we are conscious of in the broad and inclusive sense, including the successive notes, the sunlight, my surging joy, and fading anxiety – come and pass away in phenomenal time (a) as if in a dynamic flow: hence our ordinary language features the well-known phrase 'stream of consciousness' as an apt description of our conscious life. Moreover, the conscious contents all appear (b) before, or at the same time as, or after one another in phenomenal time: they all appear in a topological structure defined by the temporal 'before' and 'after' relations (as opposed to the spatial 'vertical to' and 'horizontal to' relations). I will speak of the conscious contents as appearing within *a phenomenal temporal structure*. By 'phenomenal temporal structure', I mean the apparent temporal structure in our consciousness defined by (a) the dynamic flow and (b) the temporal topology of 'before' and 'after' relations.

Indeed, as I will argue, it is with *necessity* that everything of which we are conscious appears in the phenomenal temporal structure: more precisely, *necessarily*, for a (a) unified subject who is (b) ever temporally conscious like us, all her conscious contents appear in the phenomenal temporal structure. This necessity has an important revelation: it reveals that the phenomenal temporal structure is essentially constitutive of our consciousness in the sense that it is what makes our consciousness as such possible in the first place. The phenomenal temporal structure cannot be the simple result of our perceiving objective time ‘from the outside’. Instead, I propose that the phenomenal temporal structure is what our mind imposes upon our conscious experience ‘from the inside’ as its essential constitutive precondition.

My proposal is inspired by Immanuel Kant’s (1781/1787/1996) view, according to which time is the innate mind-imposed structure of our conscious experience. On one interpretation, Kant holds that time is the *a priori* form of all possible experiences, where ‘*a priori*’ is understood as ‘not empirically perceived’ (B2) and ‘experiences’ narrowly as ‘conscious episodes’. This view contrasts with the common view that time is a feature of the world that we can and do perceive. However, if Kant and I are right, our mind does not perceive time, but instead imposes time upon our conscious experience. For this reason, we do not know whether things as they are independent of our conscious experience are really temporal in the way we experience them to be – coming and passing away as if in a dynamic flow.

I begin by presenting and clarifying the leading question around which this article is structured:

what explains the fact that necessarily all our conscious contents appear in phenomenal time? (**Section II**). To clarify what this question means, I distinguish phenomenal and objective time (**Section III**). I consider how a (higher order) theory of consciousness can be supplemented to account for the necessary ubiquity of phenomenal temporality in our consciousness. I contend that such a theory needs to posit that necessarily, to any conscious content of ours, there must correspond a temporal representation that represents this content as temporal (**Section IV**). On a *causalist proposal* that I proceed to criticize, it is purely *via* a causal mechanism – one akin to the perceptual causal mechanism involved in ordinary perception – that every conscious content of ours invariably has a corresponding representation that represents it as temporal: we invariably (causally) ‘perceive’ every conscious content of ours as temporal (**Section V**). However, it is with *necessity* that for a subject (a) who has a unified consciousness and (b) who is ever temporally conscious like us, all her conscious contents are represented as temporal (**Section VI**), yet a causal mechanism *could have failed* to represent a conscious content as temporal: we could have failed to ‘perceive’ every conscious content of ours as temporal. Therefore, I submit that phenomenal temporality cannot be the simple result of perception, but instead is an essential constitutive precondition our consciousness as such (**Sections VII and VIII**). I conclude by proposing and supporting what I call the *constitution account* for phenomenal temporality, according to which the phenomenal temporal structure is the essential mind-imposed built-in structure of our consciousness (**Sections IX, X, and XI**).

II. Clarifying the Leading Question

This article is structured around the following leading question: what explains the fact that *necessarily*, for a (a) unified subject who is (b) ever temporally conscious, all her conscious contents must appear in phenomenal time – namely, in the phenomenal temporal structure? In what follows I will sometimes use ‘us’ or ‘me’ to stand for such a subject. Thus, the leading question is: what explains the fact that it is with *necessity* that all our or my conscious contents appear in phenomenal time?

The necessary ubiquity of phenomenal temporality in our consciousness has been a key locus of investigation in the continental tradition, starting from Kant’s seminal work on the universal and necessary role the temporal structure plays in structuring sensible intuitions (Kant [1781/1787/1996]), and culminating in Edmund Husserl’s study of temporal phenomenology according to which it is the temporal structure that frames the structure of what he calls the absolute consciousness (Husserl [1893–1917/1991]). Here I follow their lead and address the question: what explains the necessary ubiquity of phenomenal temporality in our consciousness?

Three clarifications are in order. First, my question is not about how *this* particular conscious content appears before *that* particular conscious content. For example, I am not concerned with how I come to be conscious of this note C as appearing before that note D as I hear the C-D melody. Instead, what I am concerned with here is rather what explains the fact that necessarily, *all* my conscious contents appear in phenomenal time. It is this *general* question that is at stake.

Second, my question is not about the fact that all the states and processes occur within *objective*

time. Instead, it is about the distinct fact that necessarily, all my conscious contents *appear* in *phenomenal time*. It is one thing for the successive notes sounded by the brass instruments to simply occur successively in objective time. It is quite another thing for me to be conscious of them as appearing successively in phenomenal time.¹ It is *phenomenal time* that is at stake.

Third, ‘phenomenal time’ is meant to be no more ontologically committal than ‘time as consciously represented’ or ‘time as we consciously experience it’. I leave open the question of whether it is an ontologically distinct dimension over and above objective time.

III. Distinguishing Phenomenal and Objective Time

Earlier I referred to the distinction between phenomenal and objective time. It is an important distinction which merits several additional remarks. To elucidate the idea of phenomenal time, consider the following examples:

I am conscious as of [a sense of relief *after* a prolonged headache].²

I am conscious as of [the car stopping on the street *before* my friends come out of it].

I am consciousness as of [a sense of excitement surging up at the same time as one of the magnificent crescendos of Anton Bruckner’s 8th Symphony].

¹ One might think that there are conceptually possible creatures for whom objectively successive mental states are represented all at once without any temporal ordering. In their case, the phenomenal field would not be defined by a stream of conscious states featuring phenomenal temporality.

² I used ‘conscious as of’ instead of ‘conscious of’ to allow for cases of hallucination where there is no existent object. While ‘I am conscious of a pink elephant’ suggests that there is indeed a pink elephant, ‘I am conscious as of a pink elephant’ does not carry this connotation.

In these examples, I have used the brackets to mark the conscious contents – whatever I am conscious of in the broad and inclusive sense. These conscious contents extensively feature the ‘before’ and ‘after’ relations along an apparent temporal dimension. Along this dimension, various conscious contents come and pass away as if in a constant dynamic flow. I will use ‘phenomenal time’ to denote this apparent temporal dimension in our consciousness. The ‘before’ and ‘after’ of phenomenal time occur *inside* the brackets that mark conscious contents.

Phenomenal time differs from objective time. Contrast the above examples with the following:

The car stops on the street *before* my friends come out of it.

This sentence describes the temporal relation between two events in the world regardless. I will use ‘objective time’ to denote this temporal dimension of worldly states and events. The ‘before’ and ‘after’ of objective time do not occur inside the brackets that mark conscious contents.

To appreciate the distinction between phenomenal and objective time, note that from:

The car stops on the street *before* my friends come out of it.

it does not follow that:

I am conscious as of [the car stopping on the street *before* my friends come out of it].

The word ‘before’ in these two sentences designates priority in different temporal dimensions.

This point might seem obvious. However, I think we are more likely to confuse objective and phenomenal time when we speak of occurrences of mental events. To be sure, mental events appear in phenomenal time in my consciousness. However, they also occur in objective time.

Consider the following examples:

At around 3 o'clock (in objective time), I began to feel a headache.

I am conscious as of [the car stopping on the street] *before* I am conscious as of [my friends coming out of it].

I am conscious as of [a sense of relief] *after* I am conscious as of [a prolonged headache].

Occurrences of mental events are also worldly occurrences in objective time. Here 'before' and 'after' designate the 'before' and 'after' relations in objective time.

To appreciate the distinction between phenomenal and objective time, note that from:

I am conscious as of [a sense of relief] after I am conscious as of [a prolonged headache].

it does not follow that:

I am conscious as of [a sense of relief coming *after* a prolonged headache].

The reason is that it is possible that the two distinct objectively successive conscious episodes may not be represented in consciousness as subjectively successive. For instance, a person with a malfunctioning short-term memory might not recall any earlier conscious episodes at all. As William James (1890/1950) observes, a succession of experiences (i.e. conscious episodes) is not the same as an experience as of succession. In order for conscious contents to appear in phenomenal time, it is not enough that they simply occur in objective time, but I also must be conscious of them as in phenomenal time. A succession of conscious contents is not the same as a consciousness of their succession.

Let me note two differences between phenomenal and objective time. The first is what I will call the non-relativity of simultaneity. This point was made by Bertrand Russell (1914, Chapter 4), and revived by Michael Pelczar (2015, Chapter 4). According to Albert Einstein's (1905) theory of relativity, events are not absolutely simultaneous in objective time. The events that are simultaneous in one frame of reference may not be simultaneous in another frame of reference. However, in the case of phenomenal time, there is no conceptual room for the idea of temporal relativity. If an acute pain in my toe (A) appears to be simultaneous as a stroke of headache (B) within my consciousness, their phenomenal simultaneity is absolute: there is no room for a distinct frame of reference or point of view within my consciousness from which this acute pain in my toe appears to be not simultaneous with my headache. (Note that this only holds for single subjects and not across distinct subjects.) Therefore, the well-known relativity of simultaneity characteristic of objective time does not transfer over to phenomenal time.

Here lies the deep difference between phenomenal and objective time. Objective time is a dimension of the four-dimensional spacetime, which affords different points of view on two spatiotemporal events. Here the multiplicity of frames of reference allows for the relativity of simultaneity of these events. However, in the case of phenomenal time, a single subject's consciousness only affords one point of view on her phenomenal states. Unlike objective time, phenomenal time has no space associated with it to allow for multiple points of view. In virtue of this separation from space and spacetime, phenomenal time is non-relativistic.

A second difference between phenomenal and objective time is that, in special circumstances, there can be discrepancies between the objective and the phenomenal durations of the same events. Empirical studies show that traumatic events such as plane crashes, car accidents, and falls from a high cliff appear to last far longer to their survivors than they actually do (for an in-depth discussion, see Phillips [2013]). It is not merely a matter of these survivors' believing that such events last longer: they also experience them as lasting longer. For instance, a sudden fall from a high cliff lasts just a few seconds, but the survivor experiences it as lasting several times – sometimes even ten times – longer than it actually is.

The upshot so far is that one must clearly distinguish between phenomenal and objective time: they have different properties. My question concerns what explains the fact that necessarily, all our conscious contents appear in *phenomenal time*, and not how every worldly event occurs in *objective time*. Merely saying that all our conscious contents occur in *objective time* does not answer my question of what explains the necessity with which they appear in *phenomenal time*.

IV. Supplementing A (Higher Order) Theory of Consciousness

In accounting for the necessary ubiquity of phenomenal temporality in our consciousness, it is natural to begin by considering how a theory of state consciousness could be supplemented for the task. Let me begin with a popular cluster of theories which are labeled as 'higher order theories of consciousness'. They are a family of theories sharing the basic commitment that a state is conscious by virtue of being represented by a higher order representation (for an overall

review, see Gennaro [2004b]; Carruthers [2016]). Some proponents take the requisite higher order state to be a perception (Armstrong [1968], [1984]; Lycan [1987], [1996], [2004]), while others take it to be a thought (Dennett [1978]; Rosenthal [1986], [1993], [2005]; Gennaro [1992], [2002], [2004a]; Carruthers [1996], [2000], [2005]). Some require that the higher order state be actually present (Rosenthal [1986], [1993], [2005]), while others hold that the presence of a suitable disposition suffices (Dennett [1978]; Carruthers [1996], [2000], [2005]). The specific details of these variants do not concern us. What matters here is that these higher order theories share the basic commitment that a state is conscious by virtue of being represented by a higher order representation.

How can such higher-order theories be supplemented to account for the necessary ubiquity of phenomenal temporality in our consciousness? One proposal is as follows (here, as in what follows, I will occasionally use ‘conscious states’ to refer to conscious contents):

HO-0: *Necessarily*, for any conscious state S1, it – together with its corresponding higher-order representation that make it conscious – occurs in objective time.

However, this proposal is inadequate. As I remarked earlier, the fact that a series of conscious states actually occur in objective time – which is a fact about their objective temporality – is conceptually distinct from the fact that they have the phenomenal feature of appearing within an apparent dimension of time – which is a fact about their phenomenal temporality.

A second proposal is that necessarily, for any conscious lower order state, there is a higher order *temporal* representation that represents that lower order state as temporal. More explicitly:

HO: *Necessarily*, for any conscious state S1, it is correlated with a higher order temporal representation S2 that *represents S1 as temporal*.

HO avoids the problem HO-0 faces. Note that HO demands a necessary universal correlation between lower order states and higher order temporal representations. Could this demand be plausibly met? If it could, what would the underlying mental mechanism have to be like?

V. A Causalist Account for Phenomenal Temporality (to be Criticized)

One suggestion is what I will call the causalist account, according to which it is *purely* by means of some *causal* mechanism that this correlation obtains:

CT: It is *purely via* some *causal* mechanism CM that a lower order conscious state S1 is correlated with a higher order temporal representation S2 that represents S1 as temporal.

On this causalist account, I represent or ‘perceive’ a state in a similar way as I perceive this table as square or this clock as round: all this is done purely causally *via* some perceptual causal mechanism. Analogously, whenever a mental state of mine becomes conscious, a ‘perceptual’ causal mechanism invariably produces a temporal representation that represents that conscious content as temporal. Just as I perceive the table as square or the clock as round by means of a perceptual causal mechanism, I invariably represent or ‘perceive’ any conscious state as temporal by means of a similar ‘perceptual’ causal mechanism.

Note that CT is a natural extension of the causalist variant of the higher order theory of state

consciousness. The causalist variant specifies that the requisite higher order representations are causally produced, with their lineage traceable to a causal mechanism that has the lower order state as its input. This variant has prominent proponents, such as Armstrong (1968), Lycan (1996), and Rosenthal (2005). One major virtue is that it leaves no mystery about the genesis of the posited higher order states, since it is readily explicable in naturalistic terms how they are produced: we are told that it is some perceptual causal mechanism that does the work.

VI. The Necessity of Exclusive Phenomenal Temporality

CT is admittedly tempting. However, I argue that it is in fact inadequate. The reason is that CT is insufficient to explain the *necessity* with which our consciousness features phenomenal temporality. I claim that:

N: *Necessarily*, for any subject (a) who has a unified consciousness and (b) who is ever temporally conscious, all her conscious contents must appear in phenomenal time.

Necessarily, in such a subject, a state's being conscious entails its being temporally conscious: *necessarily*, such a subject is exclusively temporally conscious. This section (**Section VI**) is devoted to supporting this necessity claim. The next section (**Section VII**) spells out the reason why CT is inadequate given this claim.

Note that I do not claim that any subject whatsoever must be temporally conscious. I concede that it is possible that there are other kinds of atemporally conscious subjects, of which God is arguably a prominent example. A long and honorable tradition in Christian theology conceives

of God – and his consciousness – as eternal in the timeless sense. Illustrious proponents of this view – eternalism about God – include St. Augustine (400/1991), Boethius (596/2001), and St. Thomas Aquinas (1273/1989). I also concede that it is possible that there are other non-divine creatures who are atemporally conscious. Such subjects are not counterexamples to my claim. My claim is rather that it is impossible for a *(a) unified subject who is (b) ever temporally conscious* to be atemporally conscious: necessarily, if (a) a subject has a *unified* consciousness, and if (b) she is *ever* temporally conscious, then she must be *exclusively* temporally conscious. Paradigmatic examples of such subjects are normal human subjects like us: it is with necessity that our consciousness is temporal.

VI.i. Arguing for the Necessity of Exclusive Phenomenal Temporality

However, why think that this necessity claim is true? Let me begin by noting my profound puzzlement over the very idea of an atemporally conscious state of mine. What would it be like for me to be atemporally conscious? The issue is not merely that I have never encountered an atemporally conscious state in my consciousness: indeed I have tried to look inside my mind and have found none. The issue is that I do not understand what it would even mean for me, now temporally conscious, to come to be atemporally conscious. I do not comprehend the very idea of my being atemporally conscious.

Perhaps this suffices to convince the reader of my claim about the necessity of our exclusive phenomenal temporality. However, I think more can be said. The problem is that an atemporally

conscious state would be entirely discrete, static, durationless, and isolated. It would appear neither before, nor simultaneous with, nor after any other conscious state. For a subject (a) who has a unified consciousness and (b) who is ever temporally conscious like us, an atemporally conscious state would constitute a *distinct center of consciousness* that cannot be properly integrated with all the other conscious states of hers. (Here is an example of distinctness: your consciousness is in this sense distinct from mine.)

(i) Suppose the atemporally conscious state occurs at the same objective time as some temporally conscious states of mine. But then I cannot be co-conscious of this atemporally conscious state together with these other temporally conscious states. For otherwise, if they should appear together, what would it be like from my first-person point of view? How would the atemporally conscious state appear together with the other temporally conscious states? What would their co-appearing amount to from my first-person point of view?

Conscious states enter my *phenomenal arena*, so to speak, either side by side (e.g. headache together with stomachache) or phenomenally interpenetrated (e.g. nostalgic bittersweetness) (Bergson [1889]), or one immediately after another (e.g. new joy dissolving old anxiety). Yet if they appear side by side or phenomenally interpenetrated, they appear as simultaneous. However, if they appear one immediately after another, they appear as successive. In either case, they appear in phenomenal time. To make sense of the idea of phenomenal togetherness, we must construe it in phenomenal temporal terms: the co-appearing of conscious states could only be understood as their appearing together *within the phenomenal temporal structure*.

Therefore, if I were conscious of the atemporally conscious state and of a temporally conscious state at the same objective time, I could not be conscious of them as phenomenally together despite their objective simultaneity. In fact, I would not even have a unified consciousness, but rather two distinct centers of consciousness, each isolated from the other, violating condition (a). There would be no coherent single first-person point of view to which both states appear, even if they should occur at the same objective time. For me the atemporally conscious state would constitute a separate center of consciousness that is as distinct as yours from mine.

(ii) But what if the atemporally conscious state should occur at a different objective time from other conscious states of mine? I think the mere fact of its occurring at a different objective time would not change the fact that it remains no less a distinct center of consciousness. The situation would be essentially the same as that of someone with dissociative identity disorder (also known as the multiple personality disorder), where different subjects with distinct centers of consciousness emerge in a single body at different objective times. These subjects are as distinct as you and me. Their experiences and memories are mutually isolated. Their centers of consciousness are as distinct as yours and mine. I submit that an atemporally conscious state would constitute a likewise distinct center of consciousness. The existence of such a distinct center of consciousness would dissolve the unity of my consciousness, violating condition (a).

VI.ii. Objections and Replies

One might object that perhaps I can still remember an atemporally conscious state at another objective time, and this is enough for it to count as a state of mine. I reply that this proposal is problematic for several reasons. First, the issue is not merely that I have no recollection of such a state despite my best effort to search for one through attentive and conscientious introspection. The issue is that I do not even understand what it would be like for me to be in an atemporally conscious state, let alone to remember it.

Second, the conscious content associated with such an atemporally conscious state could not be anything temporal: it could only be an atemporal object – such as a mathematical object: $5 + 7 = 12$. However, it is far from clear how this conscious content could be properly integrated with other conscious states of mine, if I could not even be co-conscious of it together with any other conscious contents of mine – such as the various propositions about the mathematical properties and relations of numbers and other mathematical objects. In mathematical reasoning, I draw conclusions on the basis of premises according to a set of inferential rules. This requires me to be able to be conscious of the conclusion *as based on*, or *warranted by*, the premises at the time of my drawing the inference, and for this reason, I must be able to be *co-conscious* of the conclusion *together with* at least some of the premises. However, as I have argued above, co-consciousness presupposes that the states of which I am co-conscious appear together in the phenomenal temporal structure. I do not see how an atemporally conscious state – such as a putatively timeless consciousness of $5 + 7 = 12$ – can be properly integrated with any other conscious states of mine so as to count as belonging to me. Even a mathematical object such as $5 + 7 = 12$ – arguably the best candidate for an atemporally conscious content – must appear

in the phenomenal temporal structure in order for it *qua* conscious content to have any causal integration with other conscious contents of mine in mathematical reasoning. However, if it is excluded from all mathematical reasoning that I can perform, how else can such an atemporally conscious content as $5 + 7 = 12$ be properly integrated? I see no viable alternative.

One might object that perhaps some sort of causal integration could suffice even without co-consciousness for an atemporally conscious state to count as mine. Certain unconscious states, for instance unconscious perceptions, are causally integrated with other states so that they play a causal role in producing further conscious states in a subject. Empirical studies have shown that unconscious perceptions exist and do belong to a unified subject who is ever temporally conscious. Why could not the same hold for an atemporally conscious state? Why could it not be causally integrated with other temporally conscious states of such a subject in a way analogous to the way in which unconscious perceptual states are?

I reply that the analogy between such a state and the unconscious states that belong to such a subject breaks down in such a way that disqualifies the attribution of such a state to this subject, but not that of the unconscious states in question. Note that not every sort of causal integration is sufficient to underwrite proper ownership: only a particular kind is. An unconscious pain of mine and a conscious pain of yours might both causally impact my mental states and behaviors. However, the mere fact that your pain causes my perception of your state and my reaction to show concern for you is insufficient to make your pain belong to me as it belongs to you. A state can belong to a subject with a unified consciousness *only if* it can be causally integrated

in a such a way that it can causally interact with her states *via an intrapersonal* – and not merely an *interpersonal* – mechanism. Even though your pain causally impacts both you and me, it does so in very different ways, and I do not have any peculiar first personal introspective access to your pain as you do.³ My access to your pain, unlike yours, is based on external observation.

An *unconscious* state can be intrapersonally integrated, so long as its causal input lies in the domain of the factors that directly produce my mental states and behaviors. However, this can be the case *only because* it *does not* constitute a distinct and isolated center or stream of consciousness from those states of hers. Yet an atemporally conscious state *does*. Having first person access to such a state *excludes* my having first person access to other temporally conscious states of mine on pain of dissolving my unity of consciousness. An atemporally conscious state objectively simultaneous with another temporally conscious state of mine would constitute a center of consciousness as distinct from mine as yours from mine. The situation would not change simply because the atemporally conscious state were to occur at a different objective time: it would still be entirely isolated from other temporally conscious states of mine, thereby threatening my unity of consciousness, just as multiple personality disorder does to its patient. I do not see how such an atemporally conscious state could be any less interpersonally distinct from my consciousness than a conscious state of yours is. At any rate, the objector owes us a plausible account of how an atemporally conscious state could count as mine, while still preserving the asymmetry that discounts your conscious states as

³ For a characterization of the peculiar access that introspection affords us, see Byrne (2018, Chapter 1).

mine. I do not see how this could be done.

Therefore, I conclude that:

N: *Necessarily*, for any subject (a) who has a unified consciousness and (b) who is ever temporally conscious like us, all her conscious contents appear in phenomenal time.

Note that this claim is not analytic. It is not merely a matter of what ‘unified consciousness’ means that it is with necessity that the consciousness of such a subject features phenomenal temporality. For example, in speaking of a divine consciousness that embraces everything in one single atemporal intuition, Kant (1781/1787/1996) is not asserting a flat contradiction. In fact, in agreement with Kant, I do allow for a unified conscious to be atemporally conscious in the case of God. (And it goes without saying that being *ever* temporally conscious does not analytically entail being *exclusively* temporally conscious.) Therefore, the necessity claim that I propose is not merely a conceptual matter. The situation here is similar to the fact that it is with necessity that our visual experience of spatial objects features fewer than four dimensions: it is not merely a matter of what ‘our visual experience’ or ‘a unified conscious subject’ means that it is with necessity that our visual experience features fewer than four dimensions or that our consciousness features phenomenal temporality.

VII. The Non-Causal Basis of Phenomenal Temporality

Within the context of a higher order theory of consciousness, the leading question boils down

to the problem of explaining the following necessary fact:

HO: *Necessarily*, in a subject (a) who has a unified consciousness and (b) who is ever temporally conscious, for any conscious state S1, it is correlated with a temporal representation S2 that represents S1 as temporal.

Are CT and the causal mechanism it posits adequate for the task?

I think not. My reason is as follows:

[1] *Necessarily*, for a subject (a) who has a unified consciousness and (b) who is ever temporally conscious like us, all her conscious contents appear in phenomenal time.

[2] If [1] is true, then *necessarily*, in such a subject, for any conscious state S1, it is correlated with a temporal representation S2 that represents S1 as temporal.

[3] For any causal mechanism, it remains contingent that it functions properly to produce the requisite temporal representation.

[4] If [1], [2], and [3] are true, no causal mechanism *alone* is adequate to ensure that *necessarily*, in such a subject, for any conscious state S1, it is correlated with a temporal representation S2 that represents S1 as temporal.

[5] Therefore, no causal mechanism *alone* is adequate to ensure that *necessarily*, in such a subject, for any conscious state S1, it is correlated with a temporal representation S2 that represents S1 as temporal.

To appreciate the force of this argument, consider how the causalist account might try to explain

HO by revising CT. One suggestion is as follows:

CT-0: *Invariably*, in such a subject, for any conscious state S1, a causal mechanism produces a corresponding higher order temporal representation S2 that represents S1 as temporal.

However, CT-0 is inadequate. For *invariability* falls short of *necessity*: invariability of success does not rule out possibility of failure. Even if a causal mechanism invariably produces the requisite higher order temporal representation for any conscious state, due to the contingency that characterizes the operation of any causal mechanism, it still could have backfired and failed to produce that temporal representation.

An illustrative comparison might be helpful. My visual perceptual mechanism could have backfired and failed to represent the table as square or the clock as round, as in the cases of misperception, even if it has invariably succeeded in doing this so far. Similarly, the causal mechanism could have backfired and failed to represent a conscious content of mine as temporal, even if it has so far enjoyed invariable success. It is this *contingency* of success inherent in any causal mechanism that makes it inadequate to ensure the *necessity* with which every conscious content of mine appears in phenomenal time.

Therefore, what we need is a proposal stronger than CT-0. Another suggestion, which preserves the causalist spirit that motivates CT-0, is as follows:

CT-1: *Necessarily*, in such a subject, for any conscious state S1, a causal mechanism produces a higher order temporal representation S2 that represents S1 as temporal.

However, the initial problem is now pushed back to the issue of how a causal mechanism alone

could ensure that it not only *invariably* but also *necessarily* produces the higher order temporal representation S2 that represents S1 as temporal. The necessity characteristic of the correlation between any conscious state and its corresponding higher order temporal representation still remains unexplained even if an appeal is made to the operation of a causal mechanism: again, due to the general contingency that characterizes the operation of any causal mechanism, the causal mechanism could have backfired and failed to produce that temporal representation.

Therefore, mere appeals to causal mechanisms alone, such as CT and its closely related variants, are at best inadequate for our purposes. I submit that the correct account is that:

P: What makes a state feature phenomenal temporality is *essentially constitutive* of what makes that state conscious.

What this implies is that the operation of any causal mechanism alone is *insufficient* to ensure the necessity with which our consciousness features phenomenal temporality, so that in this sense, the basis of phenomenal temporality cannot be *purely* causal.

Note that my view does not rule out the possibility that a causal mechanism might still be *needed* for consciousness to feature phenomenal temporality. P by itself is neutral on whether the necessary correlation between any conscious state and its temporal representation is *in part* produced causally. Indeed, P is compatible with a causalist account such as CT-1. My point is merely that a causal mechanism alone is insufficient. I leave open the question of whether such a causal mechanism is necessary.

Indeed, one need not be committed to any causalist account at all to accept P. There is room for the view that P is true even if no causal mechanism is involved at all. Such a view would hold that no causal mechanism is even needed for consciousness to feature phenomenal temporality. I will not evaluate the causalist and the non-causalist view here. My point is that regardless of whether a causal mechanism is needed or not, in order to explain phenomenal temporality, an *additional* appeal must be made to the claim that phenomenal temporality is *essentially constitutive* of consciousness: the basis of phenomenal temporality cannot be *purely* causal.

VIII. The Argument Generalized

While I have framed my argument in the context of higher order theories of consciousness, my argument is perfectly general. Its core structure is as follows:

AG: *Necessarily*, for a subject (a) who has a unified consciousness and (b) who is ever temporally conscious like us, every conscious state of hers features phenomenal temporality. However, this necessity can be explained only by appealing to the claim that what makes a state feature phenomenal temporality is essentially constitutive of what makes this state conscious in the first place.

This result does not depend on the truth of the higher order theories: it is applicable, with the requisite changes, to other theories of consciousness as well.

IX. A Constitution Account for Phenomenal Temporality

Taking this result seriously, I now propose, in a Kantian spirit, that:

CA: Our consciousness as such has a built-in phenomenal temporal structure defined by (a) the dynamic flow and (b) the temporal topology featuring the ‘before’ and ‘after’ relations. This structure is what makes our consciousness as such possible in the first place. This structure cannot be the simple result of our perceiving objective time *via* some ‘perceptual’ causal mechanism. Instead, most plausibly, this structure has its origin in our mind: it is what our mind imposes upon our consciousness as its essential constitutive precondition.

This proposal ensures that every state of ours comes to be conscious only by virtue of featuring phenomenal temporality, thereby explaining the necessary ubiquity of phenomenal temporality in our consciousness.

The following illustrative metaphor might be helpful. Think of our consciousness as a window, and for a state to be conscious is for it to enter the window. The window has a certain structure which constitutes the window and without which the window cannot exist, and it is only by conforming to this structure that something can enter the window. I submit that the phenomenal temporal structure is the built-in structure of the window of our consciousness, and it is only by virtue of conforming to this structure that a state can enter the window of our consciousness and thereby become conscious. The structure does not come from the outside the window: it is not causally thrust in, so to speak, together with various perceptual contents – the successive notes in a melody, the sunlight sifting through the drawn shades, the cars moving on the street, etc. – through any perceptual process. Rather, the phenomenal temporal structure is the inherent

mind-imposed built-in structure of our consciousness as such.

Note that in real life the geometry of the window glass can lead to certain optical distortions of the original image. Well-known examples are the barrel and the pincushion distortions caused by curved lenses. If one looks at the undistorted image in the picture on the left (see Figure 1.1 below) through curved lenses, one sees the original image distorted in ways represented in the other two pictures. I submit that analogously, the topological structure of the temporal structure can lead to certain phenomenal distortions of objective time.⁴ First, the phenomenal temporal structure features a *non-relativistic* topology of time to which conscious states must conform, even though objective time is *relativistic* according to our best science. Second, the phenomenal temporal structure features a *dynamic flow*, even if our best science posits no such flow in objective time. On my proposal, we must see the world through the curved ‘lens’ of phenomenal time, which in some ways distorts the ‘original image’ of objective time, just as the curved lenses distort an optical image through the barrel and the pincushion distortions.

⁴ Thanks to Derk Pereboom for this helpful suggestion. See the HyperPhysics entry for more information and illustrations: < <http://hyperphysics.phy-astr.gsu.edu/hbase/geoopt/aber3.html> >.

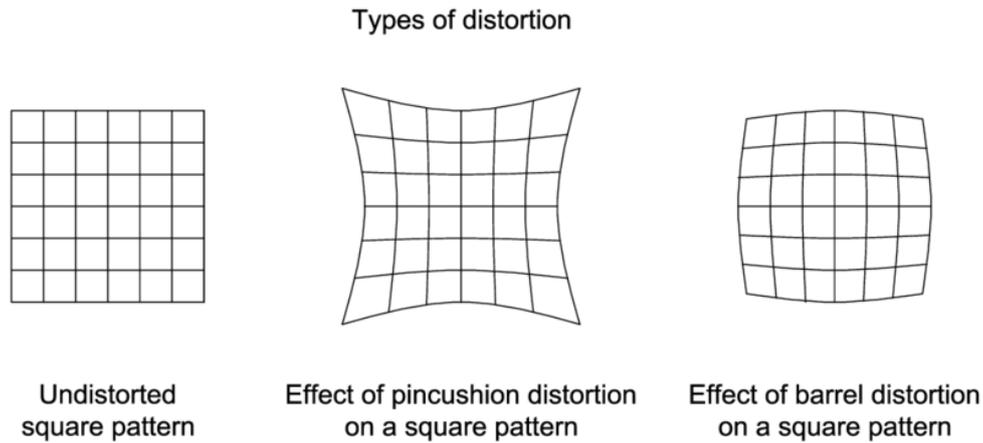


Figure 1.1

X. Clarifying the Constitution Account

On an innocent reading, my proposal fits nicely with what our best scientific theory says about objective time. Our best physics does not posit any physical counterpart to the dynamic flow that features ubiquitously in phenomenal time. Objective time is static, so to speak, in the sense that it does not flow. My constitution account provides an attractive solution that negotiates between what Wilfrid Sellars (1962) calls the manifest and the scientific image of reality. On my proposal, our phenomenology as of a dynamic flow is the way our consciousness as such is constituted. The dynamic flow of phenomenal time simply is what our mind imposes upon our conscious experience, and there is no implication that objective time itself must feature any corresponding flow. This middle position allows us to take what our best scientific theory says about objective time at face value without doing violence to our phenomenology of time.

Thus, my constitution account provides the B-theorist of time with an effective response to the

A-theorist's challenge that the B-theory is explanatorily inadequate. The A-theorist conceives of objective time as featuring a dynamic flow, and this the B-theorist denies. A common challenge to the B-theory is that it cannot easily explain how our consciousness features a phenomenal flow, whereas the A-theory can explain it simply as the result of our directly perceiving the objective flow. By an inference to the best explanation, the A-theorist declares that most plausibly, there really is an objective flow. Now, the B-theorist can appeal to my constitution account as an explanation for the ubiquity of the phenomenal flow, and then turn the A-theorist's argument on its head. If our phenomenology as of a dynamic flow in fact embodies the constitutive structure which our mind imposes upon our consciousness, then the A-theorist is no longer warranted to assert that positing a dynamic flow in objective time best explains our consciousness as of a dynamic flow.

The attractiveness of my constitution account should not be obscured by any misunderstanding. First, in saying that the *phenomenal temporal structure* cannot be the simple result of direct perception but is the mind-imposed constitutive structure of our consciousness, I do not imply that our representations of *specific temporal relations* among particular conscious states also involve no causal mechanism. There is indeed room for the claim that it is still *via* some causal mechanism involved in a perceptual process that I represent the specific temporal relations among particular conscious states of mine, with the result that, within the mind-imposed built-in temporal structure, these states end up appearing before, after, or simultaneous with those states. There is still a sense in which I can be said to *perceive* the *specific temporal relations* among particular states, even if the *phenomenal temporal structure* in which these states are

embedded cannot be the simple result of perception.

Second, my constitution account does not imply that time is purely ideal. Compatible with my account is the claim that reality in itself features objective time with some or all of the parallel properties of phenomenal time. This possibility is ignored by Kant (1781/1787/1996), who, having in effect argued for a constitution account of phenomenal temporality, jumps to the unwarranted conclusion that reality in itself features nothing that resembles phenomenal time. Critics over the years have rightly complained that this omission is a mistake. Similarly, it is a mistake to interpret my account as entailing the problematic claim that reality in itself features no properties that resemble some of the properties of phenomenal time.

However, apart from the innocent construal, there is another skeptical reading of my proposal that the phenomenal temporal structure is what our mind imposes upon our consciousness. On this skeptical reading, inspired by Kant's transcendental idealism about time (1781/1787/1996), we do not know whether things as they are independent of our conscious experience are really temporal in any sense at all. (a) Perhaps they still come and pass away one after another in an (objective) dynamic flow. Or (b) perhaps they do not feature any objective flow, but they still occur one after another in a temporal topology defined by the 'before' and 'after' relations, in accordance with our best science. Or (c) perhaps they are not temporal in any sense, featuring no flow and no temporal topology defined by the 'before' and 'after' relations at all. (It seems impossible for there to be a flow yet without any temporal topology defined by the 'before' and 'after' relations.) On the skeptical reading, it is beyond our power to know which of the three

possibilities actually obtains. At any rate, it is beyond the scope of this article to adjudicate between the two readings, and I leave this issue to the reader's choice.

Indeed, I have called my constitution account Kantian because it bears much resemblance to an interpretation of Kant's view that time is the *a priori* form of all possible experiences according to which '*a priori*' is understood as 'not empirically perceived' (1781/1787/1996, B2) and 'experiences' narrowly as 'conscious episodes'. In a Kantian spirit, I have argued that certain structural features of consciousness – such as phenomenal temporality – cannot be the simple results of the contingent operations of 'perceptual' causal mechanisms. Instead, phenomenal temporality embodies the very way in which our consciousness is constituted.

XI. Against an Alternative: No Escape from Constitution

Perhaps one might find it unsettling that my proposal – especially on the skeptical reading – shows much affinity with Kant's transcendental idealism about time. Indeed one might object by proposing an alternative account that avoids any trace of Kantian transcendental idealism while still accommodating, as my proposal does, the claim that necessarily, for a unified subject who is ever temporally conscious, all her conscious contents must appear in phenomenal time. The key idea is that it might still be through a contingent 'perceptual' causal mechanism that I come to represent every conscious content of mine as temporal. This is because perhaps my being a unified subject who is ever temporally conscious might *depend on* having such a time-

perceiving mechanism: if I should lose this mechanism, then I would cease to be unified.⁵

In reply I concede that this proposal is indeed a theoretical possibility. However, I think it faces several issues. First, it is not clear why the unity of my consciousness should depend on a contingent time-perceiving causal mechanism. It is not obvious why *time perception* should play an essential role in grounding the unity of my consciousness. Space perception does not play such a role, let alone perceptions of other sensible qualities. Moreover, it is not clear what is so special about *perception* that makes the unity of my consciousness hinge so deeply upon it. Why should one posit *perception*, and among numerous types of perception, *time perception*, as essential to, or constitutive of, my unity of consciousness?

Second, even granting for the sake of the argument that *possibly* my being a unified temporally conscious subject *might* depend on my having such a time-perceiving causal mechanism, it is not clear why *necessarily* this *must* be the case. Yet a merely contingent dependence would be insufficient to accommodate the necessity with which all my conscious contents appear in phenomenal time: there would be no explanation for this necessity in the possible worlds where the unity of my consciousness does not depend on such a time-perceiving causal mechanism. The objector would have to show that this time-perceiving mechanism is *essentially* – and not just *contingently* – involved in producing or ensuring the unity of my consciousness. However, I do not see why this *must* – and not just *might* – be the case.

⁵ Thanks to Nico Silins for this acute point.

Third, this alternative proposal does not have as many theoretical benefits as my constitution account does. For instance, this proposal does not as easily allow one to say, as my proposal does, that our phenomenology as of a dynamic flow is the result of our mental imposition, with no implication that objective time must feature a corresponding flow, in accordance with our best science today. The alternative proposal does not as easily allow, as my proposal does, for an easy reconciliation between the manifest and the scientific image of time.

Fourth, unlike objective time, phenomenal time features a non-relativistic topology. Moreover, as I will argue in the next chapter, our immediate present consciousness is not instantaneous. Rather, it has a short phenomenal duration, and more importantly, this phenomenal duration is mereologically inverted in the sense that the parts depend on the whole: any of its shorter constituent durations cannot exist except as part of the whole duration. However, objective time is not likewise mereologically inverted: rather, the whole of objective time depends on its parts. The best explanation for the non-relativistic topology and the mereological inversion of our phenomenal present – properties that distinguish phenomenal time from objective time – is that they are the results of mental imposition. At any rate, I do not see how mere perception of objective time could explain such properties which are distinctive of phenomenal time.

Finally, note that the alternative proposal is *not purely* causalist. It rests on the following claim:

WC: Time perception is essentially *constitutive* of the unity of consciousness in a unified subject who is ever temporally conscious.

This is a constitution claim, consistent with the one that I put forth earlier:

P: In such a subject what makes a state feature phenomenal temporality is *essentially constitutive* of what makes that state conscious.

Even if granting for the sake of the argument one should hesitate to accept my constitution account which posits the phenomenal temporal structure as the result of mental imposition, one must at least concede this much: it is *not purely via* a causal mechanism that the consciousness of a unified subject who is ever temporally conscious features phenomenal temporality. Instead, given its necessity, one must appeal to the notion of constitution in one form or another.

WORKS CITED

Aquinas, T. (1273/1989) T. McDermott (tr.). *Summa Theologiae: A Concise Translation*. Notre Dame: Ave Maria Press.

Armstrong, D. (1968) *A Materialist Theory of the Mind*. London: Routledge.

Armstrong, D. (1984) 'Consciousness and Causality,' in D. Armstrong and N. Malcolm (eds.), *Consciousness and Causality*, Oxford: Blackwell.

Augustine, A. (400/1991) H. Chadwick (tr.). *Confessions*. Oxford: Oxford University Press.

Bergson, H. (1889) *Time and Free Will*. F. L. Pogson (tr.). Mineola: Dover.

Boethius, A. (596/2001) J. C. Relihan (tr.). *Consolation of Philosophy*. Indianapolis and Cambridge: Hackett.

Byrne, A. (2018) *Transparency and Self-Knowledge*. Oxford: Oxford University Press.

Carruthers, P. (1996) *Language, Thought and Consciousness*. Cambridge: Cambridge University Press.

Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.

Carruthers, P. (2005) *Consciousness: Essays from a Higher-Order Perspective*. Oxford: Oxford University Press.

Carruthers, P. (2016) 'Higher Order Theories of Consciousness,' in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.

Dennett, D. (1978) *Brainstorms*. Cambridge, MA: MIT Press.

Dennett, D. (1991) *Consciousness Explained*. London: Allen Lane.

Gennaro, R. (1992) 'Consciousness, Self-Consciousness, and Episodic Memory,' *Philosophical Psychology*, 5: 333–47.

Gennaro, R. (2002) 'Jean-Paul Sartre and the HOT Theory of Consciousness,' *Canadian Journal of Philosophy*, 32: 293–330.

Gennaro, R. (2004a) 'Higher-Order Thoughts, Animal Consciousness, and Misrepresentation: A Reply to Carruthers and Levine,' in R. Gennaro (ed.), *Higher Order Theories of*

Consciousness: An Anthology, Amsterdam: John Benjamins.

Gennaro, R. (2004b) 'Introduction,' in R. Gennaro (ed.), *Higher-Order Theories of Consciousness*, Philadelphia: John Benjamins.

Husserl, E. (1893–1917/1991) *On the Phenomenology of the Consciousness of Internal Time (1893–1917)*. J. B. Brough (tr.). Dordrecht: Kluwer.

James, W. (1890/1950) *Principles of Psychology*. Mineola: Dover.

Kant, I. (1781/1787/1996) *Critique of Pure Reason*. W. Pluhar (tr.). Indianapolis and Cambridge: Hackett.

Lycan, W. (1987) *Consciousness*. Cambridge, MA: MIT Press.

Lycan, W. (1996) *Consciousness and Experience*. Cambridge, MA: MIT Press.

Lycan, W. (2004) 'The Superiority of HOP to HOT,' in R. Gennaro (ed.), *Higher-Order Theories of Consciousness*, Philadelphia: John Benjamins.

Pelczar, M. (2015) *Sensorama: A Phenomenalist Analysis of Spacetime and Its Contents*. Oxford: Oxford University Press.

Phillips, I. (2013) 'Perceiving the Passing of Time,' *Proceedings of the Aristotelian Society CXIII*.

Rosenthal, D. (1986) 'Two concepts of consciousness,' *Philosophical Studies*, 49: 329–359.

Rosenthal, D. (1993) 'Thinking that one thinks,' in Davies and Humphreys (eds.) (1993).

Rosenthal, D. (2005) *Consciousness and Mind*. Oxford: Oxford University Press.

Russell, B. (1914) *Our Knowledge of the External World*. London: George Allen and Unwin.

Sellars, W. (1962) 'Philosophy and the Scientific Image of Man,' in R. Colodny (ed.), *Science, Perception, and Reality*, Ridgeview: Humanities Press, 35–78.

van Gulick, R. (2001) 'Inward and Upward – Reflection, Introspection, and Self-Awareness,' *Philosophical Topics*, 28: 275–305.

van Gulick, R. (2004) 'Higher-Order Global States (HOGS): An Alternative Higher-Order Model of Consciousness,' in R. Gennaro (ed.), *Higher-Order Theories of Consciousness*, Amsterdam: John Benjamins.

van Gulick, R. (2006) 'Mirror, Mirror, Is That All,' in U. Kriegel and K. Williford (eds.), *Self-Representational Approaches to Consciousness*, Cambridge, MA: MIT Press.

CHAPTER II

Specious Present, Phenomenal Extension, and Mereological Inversion:

A Problem for Physicalism about the Mind

I. Introduction and Map of the Chapter

On my computer runs Andrei Tarkovsky's movie *Andrei Rublev*. In the church sacked by the Tartars, Rublev, the renowned Russian iconographer, tormented by the evils of the world, is confiding to his teacher Theophanes the Greek. Snow is falling through the dome of the church against the background of the scorched murals. As I am watching this scene, I hear the gentle sounds of the air conditioner, feel the coolness of the air, and experience the residual bittersweet taste of my last sip of coffee. All these goings-on appear to me within my present consciousness, i.e. within the specious present (James [1890/1950]). The specious present is the immediate phenomenal temporal structure within which various occurrent states and events appear to me. Note that the notion of the specious present pertains to the representational content of my immediate experience and not to its objective duration: the specious present is the phenomenal temporal structure framing the representational content of my immediate experience.⁶

What follows is a study of the mereological structure of the specious present and what it reveals

⁶ I will assume that every event in our experience is represented as occurring in time. But this assumption can be challenged. Thanks to an anonymous reviewer for raising this point. This is an interesting and complex issue for an independent debate which unfortunately I cannot delve into here.

about the nature of the mind. I argue that the specious present has certain features that cannot be easily explained within the framework of physicalism – the view that consciousness is nothing over and above what is physical. I contend that the specious present is a phenomenally extended whole whose parts are conceptual abstractions and are therefore dependent on this whole. However, unlike the specious present, what is physical does not have this special property of the parts being dependent on the whole. Therefore, it is unclear how the specious present can be explained in physicalistic terms.

Below I proceed as follows. I begin by providing a characterization of the specious present: it has a phenomenal temporal extension (**Section II**). I defend an existent argument for this claim against Simon Prosser’s (2018) critique, and raise some further considerations to corroborate this result (**Section III**). I then argue that the specious present is mereologically inverted, i.e. with the parts dependent on the whole rather than *vice versa*, on the basis that any of its shorter constituents cannot be conceived in the appropriate sense except as part of the phenomenally extended whole of the specious present (**Section IV**). Finally, I raise a challenge to physicalism about the mind: the mereologically inverted specious present cannot be identical to, or purely constituted by, any physical particular, which is not mereologically inverted (**Section V**).

II. Characterizing the Specious Present

The specious present is the phenomenal temporal structure of the representational content of my immediate consciousness. Two features of the specious present are worth noticing. First,

the specious present has a non-relativistic temporal topology. In the specious present, all states and events appear to me in temporal relations of priority, simultaneity, and posterity. Distinctive of this phenomenal temporal topology is the non-relativity of simultaneity. This point was made by Bertrand Russell (1914: Chapter 4), and revived by Michael Pelczar (2015: Chapter 4). According to Albert Einstein's (1905) theory of relativity, events are not absolutely simultaneous in objective time. The events that are simultaneous in one frame of reference may not be simultaneous in another frame of reference. However, in the case of phenomenal time, there is no conceptual room for the idea of temporal relativity. If an acute pain in my toe (A) appears to be simultaneous as a stroke of headache (B) within my consciousness, their phenomenal simultaneity is absolute: there is no room for a distinct frame of reference or point of view within my consciousness from which this acute pain in my toe appears to be not simultaneous with my headache. (Note that this only holds for single subjects and not across distinct subjects.) Therefore, the well-known relativity of simultaneity characteristic of objective time does not transfer over to phenomenal time.

Here lies the deep difference between phenomenal and objective time. Objective time is a dimension of the four-dimensional spacetime, which affords different points of view on two spatiotemporal events. Here the multiplicity of frames of reference allows for the relativity of simultaneity of these events. However, in the case of phenomenal time, a single subject's consciousness only affords one point of view on her phenomenal states. Unlike objective time, phenomenal time has no space associated with it to allow for multiple points of view. In virtue of this separation from space and spacetime, phenomenal time is non-relativistic.

Second, the specious present is phenomenally extended in the sense that it has a non-punctual, albeit very short, phenomenal duration. In a famous passage, William James (1890/1950) characterizes the specious present as having

no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched [...] The unit of composition of our perception of time [i.e. the specious present] is a duration [...] It is only as parts of this duration-block that the relation of succession of one end to the other is perceived. We do not first feel one end and then feel the other after it, and from the perception of the succession infer an interval of time between, but we seem to feel the interval of time as a whole, with its two ends embedded in it (James 1890/1950: 609-610).

This is a plausible phenomenological observation. James further contends that

[The duration-less present moment is] an altogether ideal abstraction, not only never realized in sense, but probably never even conceived of by those unaccustomed to philosophic meditation (James 1890/1950: 608).

His point here is that the specious present has an extended phenomenal duration, and the present deprived of any duration is a mere mathematical or philosophical abstraction to which nothing in my phenomenal consciousness corresponds.

There are two issues with James' view. Consider the more complete version of the earlier quote:

The practically cognized present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions

into time. The unit of composition of our perception of time is a duration, with a bow and a stern, as it were – a rearward – and a forward-looking end (ibid.).

The first issue concerns what James means by the phrase ‘look in two directions into time’. The straightforward reading is that he thinks that our awareness extends both into the past and into the future. This reading has James endorsing the implausible claim that we have some ‘mysterious clairvoyance’ into the future, as Barry Dainton (2011) aptly puts it. Following most interpreters, such as Dainton (2011) and Jenann Ismael (2011), I read James as saying that although I am at once aware of what has happened a short while ago up until now, I merely have some anticipation of what might come, without any mysterious awareness of the future.

The second issue is that, as Dainton (2011) points out, James seems to overestimate the length of the specious present. Here is James on the temporal span of the specious present.

Our maximum distinct intuition of duration hardly covers more than a dozen seconds (while our maximum vague intuition is probably not more than a minute or so) (James 1890/1950: 630).

Intuitively, my specious present has a fairly short temporal span: it most likely lasts no more than a second. To suggest that my immediate awareness extends ‘a dozen seconds’ into the past is not credible. Perhaps what James intended to pick out with the phrase ‘specious present’ is not exactly the same as what most of his contemporary followers claim to pick out with the same phrase.⁷ However, I will be using this term in the sense at stake in the contemporary debate, that is, to pick out the temporal structure of our immediate experience, even if that is

⁷ Thanks to an anonymous reviewer for raising this issue.

not what James originally intended. Yet the Jamesian insight remains plausible: the specious present is not instantaneous but phenomenally extended.

III. Arguing for the Phenomenally Extended Specious Present

There is a continuing debate about whether the present consciousness really has a phenomenal temporal extension. Following the seminal discussion by Edmund Husserl (1893 – 1917/1991) and James (1890/1950), many analytic philosophers, such as C. D. Broad (1923), Clement Mundle (1966), John Foster (1982), Dainton (2000), Michael Tye (2003), Matthew Soteriou (2010), Ian Phillips (2010), Oliver Rashbrook (2012), and Christoph Hoerl (2013), accept that our present experience is phenomenally extended. However, this doctrine is not without detractors, which include Christof Koch (2004), Robin Le Poidevin (2007), Philippe Chuard (2011), Valtteri Arstila (2017), and Prosser (2018) all of whom, except Prosser, subscribe to some version of the ‘snapshot’ view, on which the present experience instead is punctual, with no extended phenomenal duration. Prosser is not a snapshot theorist, although he thinks there is much to be said for the snapshot theory. (His considered view is Dennettian: he maintains that there is no matter of fact as to whether we have a phenomenally extended specious present.)

I maintain that the present experience is indeed phenomenally extended. In defense of this claim, I first explain and vindicate an existent argument against Prosser’s (2018) critique (**Section III.i**).⁸ I then raise some further considerations in favor of a phenomenally extended

⁸ For a recent extended critique of Prosser (2018), see Shardlow (2019).

specious present on the basis of the non-relativistic topology of phenomenal time (**Section III.ii**).

III.i. Vindicating an Existent Argument

A common argument for a phenomenally extended specious present appeals to the idea that in order to have experience as of motion – which we do in fact have – we must have phenomenally extended experience as of the moving object existing in different places at different times:

[1] We have experience as of motion.

[2] If [1] is true, then we must have experience as of the moving object existing in different places at different times.

[3] If [1] and [2] is true, then we must have experience with phenomenal temporal extension.

[4] Therefore, we do have experience with phenomenal temporal extension.

For instance, to experience the motion of a car we must experience the car as being at different locations at different times, and this would be impossible unless our experience had a phenomenally extended temporal content.

Although the argument appeals to our experience as of motion, the idea here is perfectly general. Motion is merely a particular kind of change: change in spatial location. We might replace motion with any other experienced change in any sense modality and in introspection.

[1*] We have experience as of *change*.

[2*] If [1*] is true, then we must have experience as of *the changing object existing in different states at different times.*

Etc.

This version of the argument may overstate what is needed for having experience as of change. What is needed is only that we have experience as of *there being* different states at different times, without the need to have any experience as of which object is changing or even as of any object changing. It seems that I might have experience as of

(a) A rise in the volume of a sound, or

(b) A fall in temperature, or

(c) An increase in the intensity of malodor, or

(d) An emergence of a new idea, coupled with a sudden rise of a sense of hope,

without thereby being aware of which object is changing or as of any object changing. If so, we might rephrase [2*] as follows:

[2**] If [1*] is true, then we must have experience as of *there being different states at different times.*

Prosser (2018) mounts an extended criticism of this argument. He distinguishes between two kinds of change:

(i) Continuous motion and motion-like change: ‘motion and motion-like change where an instantaneous finite rate of change can be defined’ (Prosser 2018: 127), and

(ii) Discontinuous change: ‘change in which an object is in one state at all times up to and including [a certain time] t, and in a different state at all times thereafter’ (ibid.).

He then argues that for either (i) or (ii) there is no convincing argument for the phenomenally extended specious present.

Let me note that it is not entirely clear to me what Prosser means by 'motion-like change'. I think it is reasonable to surmise that this category includes continuous changes in length and size of spatial objects such as the gradual contraction of a tire and the rapid expansion of a balloon. However, it is not immediately apparent whether, and in which respects, certain continuous changes are motion-like. Consider the following:

- (a*) A quick rise in the volume of a sound;
- (b*) A rapid drop in environment temperature;
- (c*) A fast increase in the intensity of a malodor;
- (d*) A rapid magnification of a sense of hope.

Note that the changes in (a*) through (d*) are continuous in the sense that there is a way in which a rate of change can be intelligibly defined. There is no doubt that they do not fall into the category of discontinuous changes as defined in (ii). However, do they count as continuous motion-like changes as defined in (i)? In what sense are they motion-like? What exactly is motion-likeness? Prosser is not explicit on these points.

(i) Insofar as continuous motion and motion-like change are concerned, Prosser argues that [2] in the first version of the argument is false. Following Le Poidevin (2007), Prosser argues that we can have experience as of change even if we have no experience as of the object existing in different locations at different times. Prosser appeals to the waterfall illusion:

In fact empirical science has uncovered many examples of experiences of motion where there is no experience of a change of position. The best known examples of this are motion after-effect illusions such as the Waterfall Illusion. In the Waterfall Illusion the subject first looks at a scene in which there is constant motion, such as the movement of water in a waterfall. When the subject then looks at a stationary scene, illusory motion is experienced in the opposite direction to that of the genuine motion in the previous scene. Yet – at least arguably – it does not appear to the subject that there is any change in the locations of the objects in the perceived scene (Wohlgemuth 1911) (Prosser 2018: 123).

Prosser goes on to explain such phenomena as follows:

It is now widely believed that the brain computes motion at least in part using a low-level detection system, though higher level processes may also be involved. We should take it that there is some degree of independence between the computation of motion and the computation of position; normally the two harmonize, and they may interact in certain ways, but in the case of certain illusions they can come apart. (Prosser 2018: 125).

The key idea here is that there is a distinction between the low-level mechanism for motion detection and the high-level mechanism for position computation, and it is possible for the former to function independently of the latter, in which case we would have experience as of motion without experience as of the object existing in different places at different times.

In reply, let me make two points. First, the sort of motion experienced in the waterfall illusion

is in a crucial way (a) non-robust and (b) atypical. (a) It is non-robust insofar as it is phenomenally distinct from experience as of full-fledged dynamic movements involving different positions. Camden McKenna (2020: 834-835) makes a similar point: in the waterfall illusion, we do not experience the cliff face as really moving in a full-fledged sense, e.g. in the same way we experience the waterfall as moving. In fact, the majority of our experiences as of motion are unlike the waterfall illusion experience. I will call such experiences ‘experiences as of robust motion’, since there is a more convincing and genuine sense of dynamic movement than that found in the waterfall illusion. Since the operation of the mechanism for motion computation alone is inadequate to convey a convincing sense of dynamic movement, it is plausible that our experience as of robust motion is produced by the cooperation of the mechanism for motion computation and that for position computation.

Moreover, (b) the waterfall-illusion-style experience as of non-robust motion is also atypical insofar as it only occurs in rather special circumstances. As Prosser admits, both the mechanism for motion computation and that for position computation *normally* work together in harmony (Prosser 2018: 125). Note that although the mechanism for position computation is indeed high-level relative to that for motion computation, they are both mechanisms which generate experience. The proper conclusion to draw here seems to be that in normal circumstances, the two mechanisms interact to produce full-fledged experience as of robust motion. If instead the function of the high-level mechanism for position computation were insulated from experience and produced thoughts and judgments only, then the apparently distinct locations and their spatial relations would never be experienced but merely thought or discursively grasped. Yet

this is likely not the case. We may redeem [1] by revising it as follows:

[1**] *Typically, we have full-fledged experience as of robust motion.*

Such experience plausibly requires experience as of the object existing in different places at different times.

Yet the revised argument is not conclusive. Prosser might reject [1**] by saying that in cases where we think we experience robust motion, our present experience remains instantaneous, but it is this instantaneous experience as of the object moving (*via* the low-level mechanism for motion detection) combined with our *memory* as of its distinct locations (*via* the high-level mechanism for position computation) that yields a sense of robustness.⁹ I concede that this possibility has not been ruled out, but other considerations seem to weigh against his position.

Second, even if [1**] were rejected, Prosser's objections would not block the argument for the phenomenally extended specious present, because the empirical data he invokes are restricted to visual motion detection, while the argument can appeal to other sense modalities. Recall the changes in cases (a*) through (d*). Since their detection is carried out in non-visual sense modalities, it is a logical leap to assume that these change detection processes involve a structure analogous to the visual one. I am not aware of any illusion structurally parallel to the waterfall illusion in other non-visual sense modalities. Moreover, even granting for the sake of argument that future empirical studies should reveal that these non-visual sense modalities do employ two distinct mechanisms, one for change detection and the other for identification of

⁹ Thanks to an anonymous reviewer for raising this point.

states before and after the change, it could be argued that

[1***] *Typically, we have full-fledged experience as of robust change.*

Such experience plausibly requires phenomenally extended content as of there being different states at different times. The argument for the specious present is thus vindicated.

Perhaps Prosser is best understood as offering an account for our motion perception only, and not for the other types of experiences as of change. Yet the notion of specious present at stake here denotes our immediate present consciousness, which need not be exclusively motion-perceptual. As Jack Shardlow (2020) remarks, the original motivation in James (1890/1950) for a phenomenally extended specious present does not derive only from considerations of motion perception. To block the argument from change, which considers our consciousness as of other types of changes, more needs to be said than is explicit in Prosser's account for motion perception, although I concede that I have not ruled out his account in a significant local case. More importantly, a major aim of this article is to argue against physicalism about the mind (**Section IV** and **V**). To raise this argument, defending a phenomenally extended episode of consciousness, not necessarily restricted to motion perception, will suffice for my purpose.

(ii) The second category of change Prosser considers is discontinuous change from one state to another. According to the argument he criticizes, if the content of our experience were always restricted to a singular durationless instant, then we would only experience a single unchanging state, but never a discrete jump from one state to another, which we nonetheless do experience.

To undercut the force of this argument, Prosser appeals to an empirical account for change

detection on the basis of scientific studies of change blindness. The key idea of this account is that we detect sudden changes by means of a signal known as ‘the visual transient’ which contains information that such a change has occurred but not of what has changed.

[A visual transient] is a rapid change in the luminance or color of the retinal image, of the kind that occurs when something that is visually perceived suddenly changes. Visual transients are detected in the very early stages of the visual system, and they automatically attract attention to the location at which they occur. A transient can therefore be thought of as a signal that a change has occurred at a specified location, but that carries no information about what it was that changed.

Armed with such an account, Prosser proposes a hypothesis for our detection of discontinuous changes.

Insofar as there is a genuine phenomenology associated with such changes, it is due to the content associated with the transient. This would be something like: ‘a change has just taken place here’ (where ‘here’ is substituted by some specification of a location). When one’s attention is directed to the location of the transient one can then become aware of what has just changed, but this involves a comparison of current experience with the content of a very short-term memory, rather than being an element of the immediate phenomenology. Thus, strictly speaking, for discontinuous changes there is a distinct phenomenology indicating that something has just changed, but no phenomenology indicating what has just changed. (Prosser 2018: 131).

On this view, since there is no need to posit experience as of what has changed, there is also no need to posit experience as of there being different states at different times. Thus Prosser is in

a position to deny [2**].

In reply, let me make three points. First, the empirical account for change-blindness on which Prosser's hypothesis relies does not by itself entail that we do not have experience as of there being different discrete states at different times. In fact, this account is compatible with both the claim that we do have such experience and the claim that we do not. The account itself does not specify what computational role the informational content associated with the visual transient plays in relation to the computational processes that generate experience in typical circumstances. It would be entirely consistent with this account to suppose, contrary to Prosser's hypothesis, that such informational content is also used in attention-involving computational processes that generate experience as of there being different states at different times. I am not suggesting that Prosser needs to rule out the opposing hypothesis, but merely making the weaker point that given the neutrality of the empirical account on this issue, Prosser's hypothesis remains speculative.

Second, our phenomenology would seem to weigh against Prosser's hypothesis. It is with almost equal phenomenal immediacy and vivacity that we represent the different states in a discontinuous change and those in a continuous yet robust change. If phenomenal immediacy and vivacity of this kind warrant an inference to experience as of the different states in a continuous robust change, as it indeed does to my mind, then there is no reason to reject a parallel inference to experience as of the different states in a discontinuous change.

Prosser is skeptical that we have a substantive phenomenology as of discontinuous changes apart from continuous ones (2018: 128-129). Yet, more plausibly, we do. Prosser concedes that some changes, such as a sudden change of color between red and green, cannot easily be assimilated to a continuous one, for it seems implausible to postulate any intermediate state between a red-state and a green-state (*ibid.*). Let me add that in order to discount this case, what needs to be postulated must be an infinity of intermediate states constituting a continuum, none of which I can introspect despite my most conscientious effort in ideal conditions, and this is highly implausible.

Prosser appeals to his visual transient account to capture this phenomenology. Yet his account may be insufficiently general. It is implausible that our sensory faculties are so supremely fine-grained as to be sensitive to all such putative intermediate phenomenal states, which I cannot introspect despite my best effort. By contrast, our sensory faculties most likely have minimal thresholds of sensitivity. My eyes are only sensitive to light of a certain intensity. If so, any change from a zero state (where the sensory faculty receives no input and contributes no output) to a minimal non-zero state (where the faculty is just barely activated and contributes a minimal output) is discontinuous. When I open my eyes in a room where everything in it remains changeless and is just barely lit and visible at the threshold of my visual sensitivity, my experience registers a discontinuous change from a state of darkness to a view in minimally perceivable light. It would seem that Prosser must bite the bullet against phenomenology of this kind. More importantly, it is not clear how his visual transient account can explain changes from zero states to non-zero states in such cases where nothing perceived changes.

Third, according to Prosser's hypothesis, it is by means of a comparison of the informational content of the previous state in short-term memory with that of the current experienced state that we are to be aware of what has changed in a sudden discontinuous change. What exactly is involved in this comparison, and how does it work? One answer is that the experiential content associated with the previous state is retained in short-term episodic memory, and at the time of the comparison, it is combined with the experiential content associated with the current experienced state to form an encompassing content which is extended. Such a situation is better described as one in which the short-term memory cooperates with the current experience to generate an experience with the extended content as of what has changed in a discontinuous change. This result turns out to support the argument Prosser is criticizing. However, if this answer is incorrect, then it is unclear what the nature of such a comparison is supposed to be. If it is not post-experiential, then the previous answer seems inevitable. If it is post-experiential, then it could not account for our immediate and vivid phenomenology, since on this option there would be no experiential content encompassing both the previous and the current state.

III.ii. Further Considerations in Favor of a Phenomenally Extended Specious Present

The above discussion has vindicated the existent argument in favor of a phenomenally extended specious present. To corroborate this result, I will raise some further considerations in its favor. Earlier I set out the absolute topology of phenomenal time, characterized by non-relativistic temporal relations of priority, simultaneity, and posterity. Here I submit that (a) such a peculiar

topology most plausibly already features in the experiential content, and (b) any instantaneous experience leaves no room for such a topological structure to feature in its content.

(a) The non-relativistic topology of phenomenal time most plausibly already features in the experiential content. From introspection I acquire the belief that my experience features a non-relativistic phenomenal time. Absent any good reason to think otherwise, this belief is a strong reason that my experience is indeed so. After all, in saying that phenomenal simultaneity is non-relativistic, Russell (1914) and Pelczar (2015), as I understand them, are attempting to describe what the phenomenal temporal aspect of our experience is like. They assume, quite reasonably, an inference from the introspectively acquired belief that our experience features a non-relativistic phenomenal time to the claim that our experience is indeed so.

However, is there any good reason to think otherwise? One might object that this belief could be explained away without positing any experience featuring a non-relativistic phenomenal time. On this view, our experience is confined to a singular moment which is structurally simple, and it is through the interplay between such an instantaneous experience and our memory of past instantaneous experiences that we come to have the belief that our experience features a non-relativistic phenomenal time (e.g. Dennett [1991], Chuard [2011]). However, this belief is strictly speaking false, since we do not really have such an experience.

Thus we are systematically deluded about the basic structure of my experience, or so this view suggests. This should already raise some suspicion. However, I concede that our folk beliefs

are not always veridical, and this might be a case in which they are systematically erroneous.¹⁰

Yet I think more can be said against this view. How is the interplay between the current instantaneous experience and my memory of past instantaneous experiences supposed to work? The mere sum of a currently experienced moment and remembered moments in the past yields no topology at all. More generally, a mere sum of individual instantaneous experiences is *topologically structureless*. Such a sum by itself does *not* specify that the moments belong to the *same sequence* of time by standing in (i) the *same topological structure* of time, let alone in (ii) a *non-relativistic topological structure* of time. Call this *the unity problem* for the instantaneous present view.

It might be the case that our mind imposes the relevant temporal topological structure upon our experience.¹¹ Yet we should distinguish between two kinds of imposition. The first results in the topological structure featuring in the content of the specious present itself. On this option, for reasons given below in (b), the specious present must be phenomenally extended. By contrast, the second kind of imposition generates a *mere belief* that experienced states and events relate to one another in accordance with this structure, *without* this structure featuring in the experiential content. However, it is not very plausible that this second kind of imposition actually occurs, and even if it does, it likely does so only sporadically.

¹⁰ Thanks to an anonymous reviewer for pressing this point. For more discussion on whether our folk beliefs are reliable, see Hershfield and Maglio (2020). For an illusionist view, see Miller et al. (2018). Thanks to the same reviewer for bringing my attention to these recent works. Because of its complexity, this issue, albeit interesting, demands an independent study that I unfortunately cannot carry out here.

¹¹ Thanks to an anonymous reviewer for raising this point.

For among our representations associated with phenomenal time there is a prevalent kind that is *cognitively impenetrable vis a vis* its topology. Even though we know (or strongly believe) that objective time is in fact relativistic in light of Einstein's theory, there is still a sense in which time seems non-relativistic to us: time still phenomenally appears to us as absolute. This fact evokes Russell's important insight, set out earlier, that phenomenal simultaneity is non-relativistic: there is no conceptual room for the idea of temporal relativity in our consciousness. The fact that phenomenal simultaneity remains non-relativistic despite my knowledge of (or strong belief in) Einstein's theory of relativity constitutes an instance of *cognitive impenetrability*. This cognitive impenetrability makes it plausible that it is in the content of our experience that the non-relativistic topology of phenomenal time features.

Moreover, if it were the second kind of mental imposition always at work, there would have to be some mechanism that inclines me to believe that phenomenal time features a non-relativistic topology, and such a mechanism would have to be *post-experiential*. However, it is very difficult to see what it could be in the absence of any experience as of such a topology. It is natural to appeal to memory as the mechanism at issue, but the unity problem I referenced earlier remains. My immediate experience of the current instant and my memory of another instant or series of instants do not specify that (i) these instants belong to the same unified timeline or that (ii) this unified timeline exemplifies a non-relativistic topology. However, without positing an experience as of this common topological structure, it is not clear what would account for our beliefs (i) and (ii).

(b) However, it is impossible for the non-relativistic topology of phenomenal time to feature in an instantaneous experience. The structureless point-like experiential content leaves no room for a structurally complex topological structure to feature in it: an instant with no extension is too structurally simple and extensionally limited to accommodate a constitutionally complex topological structure defined by non-relativistic relations of priority, simultaneity, and posterity.

Note that an instant *underdetermines* the topological structure of time to which it belongs. Consider the spatial parallel. A spatial point underdetermines the topology of its space. A spatial point can belong to a three-dimensional structure, e.g. Euclidean space. It can exist in a structure with more than three dimensions. It can belong to a two-dimensional structure, e.g. a plane, or a one-dimensional structure, e.g. a line. Indeed, it can even exist by itself and not as part of any higher dimensional structure of space. Now, if my field of consciousness had always been confined to a mere point without any spatial extension, it is difficult to see how the particular topological structure of three-dimensional space, in contrast to any other topology, could be determined in the experiential content.

Generally speaking, an experience restricted to a mere point-like element – a spatial point or a temporal instant – underdetermines the topological structure to which that element belongs. Consider phenomenal time. The topology of phenomenal time is that of absolute time. However, this is not the only possible temporal topology: a relativistic topology characteristic of Einsteinian spacetime is another possibility. Therefore, a temporal instant underdetermines the

topology of time to which it belongs: a mere summation of instantaneous experiences on their own falls short of yielding the particular topology of absolute time, in contrast to any other possible temporal topology. The peculiar non-relativistic topology of phenomenal time must already be given *together with* any phenomenal experience of ours in order for phenomenal time to be the way it actually is.

The core idea here is that the specific non-relativistic topology of phenomenal time supports phenomenal temporal holism *vis a vis* the specious present. Given its peculiar topology, phenomenal time is more than a mere summation of phenomenal instants on their own. A non-instantaneous experience is required to accommodate such a topology. Therefore, specious present is most plausibly phenomenally extended.

IV. Mereological Inversion of the Specious Present

Having argued that the specious present is a phenomenally extended whole, I will now argue that it is mereologically inverted in the following sense:

MI: An object O is mereologically inverted just in case the whole of O is metaphysically prior to any parts of O.

The whole is metaphysically prior to its parts in the sense that the parts depend on the whole so that they cannot exist except as parts of the whole. Now I will argue that the specious present as a phenomenal whole is metaphysically prior in this sense to any of its constituent phenomenal durations.

MI is a metaphysical claim about the structure of an object. The idea of MI comes from Immanuel Kant (1781/1787/1996), who thinks that both space and time are mereologically inverted. In his discussion of space in the transcendental aesthetic, Kant says that:

Nor [...] can these parts [of space] precede the one all-encompassing space, as its constituents, as it were (from which it can be assembled); rather, they can be thought only in it. Space is essentially one; the manifold in it, and hence the universal concept of spaces as such, rests solely on [our bringing in] limitations (Kant 1781/1787/1996: A25/B39; my underscoring).

On his view, the whole of space is metaphysically prior to its parts, because parts of space can be conceived only by imposing limitations on the whole of space. In the case of time, he makes several parallel claims:

To say that time is infinite means nothing more than that any determinate magnitude of time is possible only through limitations [put] on a single underlying time. Hence the original representation of time must be given as unlimited. But if something is such that its parts themselves and any magnitude of an object in it can be represented determinately only through limitation, then the whole representation of it cannot be given through concepts (for they contain only partial representations), but any such representation must be based on direct intuition (Kant 1781/1787/1996: A32/B48; my underscoring).

In these passages, Kant in effect proposes a test for mereological inversion by specifying a

conceivability criterion which, if met by an object, indicates that this object is mereologically inverted:

MIT: For any two objects P and O, if P is a part of O, and if P cannot be conceived in the appropriate sense except by virtue of conceiving of O as divided into parts of which P is one, then there is strong evidence that O is metaphysically prior to P, and so by definition, mereologically inverted.

Note that the sort of conceivability attempt MIT has to do with is not the attempt of conceiving P as part of O. Rather, MIT has to do with the attempt of conceiving of P *simpliciter*.

What is the appropriate sense of inconceivability here? Kant does not specify. However, in his support, I think a viable amendment of MIT is as follows:

MIT*: For any two objects P and O, if P is a part of O, and if however attentive I am and however better I am informed and capable of reasoning and imagination when I try to conceive of P, P cannot be conceived except by virtue of conceiving of O as divided into parts of which P is one, then there is strong evidence that O is metaphysically prior to P, and so by definition, mereologically inverted.

The underscored part in MIT* spells out the appropriate sense of inconceivability in MIT.

MIT* rests on the plausible idea that if it is impossible to conceive in the appropriate sense of P without conceiving of O as divided into parts of which P is one, then it is impossible that P could exist without being a part of O, and therefore, O must be metaphysically prior to P. The qualification of inconceivability ‘in the appropriate sense’ is important. It is well known that

not every sort of inconceivability is a reliable guide to impossibility. (a) I cannot conceive of a polygon with a million sides, but my inability does not entail that it cannot possibly exist. (b) A medieval alchemist claims that he cannot conceive of water as made up of more fundamental elements, but his inability does not entail that water cannot possibly be made up of more fundamental elements. However, the inconceivability invoked in examples (a) and (b) is not the one in the appropriate sense. (a) My inability to conceive of a polygon with a million sides results from my limited imagination. However, with my capacity of imagination properly augmented, there is in principle nothing to prevent me from conceiving of this polygon. (b) The medieval alchemist's proclaimed inability to conceive of water as made up of more fundamental elements results, at least in part, from his lack of knowledge of our post-medieval science. However, if, as a competent reasoner, the alchemist is informed of the discovery that water is made up H₂O molecules with H and O atoms, there is in principle nothing to prevent him from conceiving of water as made up of more fundamental elements.

Inconceivability in the appropriate sense is a strong indicator of metaphysical impossibility. I cannot appropriately conceive of a slope-less mountain, and this is strong evidence that it is metaphysically impossible. Nor can I appropriately conceive of two actual physical objects in the world as existing within two isolated spaces and bearing no spatial relations with each other, and this is strong evidence that it is metaphysically impossible for them to be so. Similarly, my inability to conceive of P in the appropriate sense except as part of O is strong evidence that it is impossible for P to exist except as part of O, and therefore, O is metaphysically prior to P.

Yet however attentive I am, and however better I am informed and capable of reasoning and imagination, for any phenomenal duration shorter than the specious present, I can conceive of it only by way of abstraction from the phenomenal background of the specious present in which it is embedded. To do so, I must first conceive of the specious present which is temporally extended, and then ideally subtract some duration from it to form a conception of a constituent of it whose duration is even shorter. The situation does not change even if my knowledge and relevant capacities are augmented to the ideal extent. The reason is that the specious present is the shortest phenomenal duration in immediate consciousness, and the act of conceiving of any shorter constituent is essentially a process of abstraction from a unified phenomenal whole. By MIT*, there is strong evidence that the specious present is mereologically inverted.

One might object that one can indeed conceive of a shorter duration on its own. For instance, when two different sounds are heard in quick succession within the duration of the specious present, it might seem that one can conceive of one of the sounds on its own.¹² I concede that one can indeed do so, but it seems to me that this can only be done by means of abstraction and idealization. I do not claim that this shorter duration is inconceivable on its own, but rather that it is inconceivable without our conceiving of the whole and then ideally subtracting from it.

Earlier I claimed that the specious present is the shortest conscious interval of our immediate experience. Yet one might challenge this claim. For even if there is a phenomenally extended specious present construed as the temporal structure of our immediate consciousness, it does

¹² Thanks to an anonymous reviewer for raising this point.

not follow that it is our minimal conscious interval. Compare it with our phenomenal space. It is far from clear that our visual field of space is the minimal phenomenal extension in our immediate experience: arguably by squinting we can diminish our visual field of space. Why think our specious present – the temporal analogue of our visual field of space – is minimal as I claim it to be?¹³ My reply is twofold. First, if a putative specious present were to contain a shorter conscious interval capable of existing on its own, then there would be no reason to refuse the shorter interval the proper title of the specious present. But then the original longer interval could not count as the specious present *in the strict sense*. Admittedly, the term ‘specious present’ has multiple meanings and referents in the history, and many philosophers, including James, do not think of it as minimal in my sense.¹⁴ Yet for my purpose, I intend the term in the strict sense. The upshot of my argument is that the specious present in this sense is (a) phenomenally extended and (b) mereologically inverted. (Note that the strict notion of the specious present is not the same as the strict mathematical notion of the present moment, which James rightly takes to be a mere abstraction from the specious present strictly construed.)

Second, as Soteriou (2013: Chapter 5) correctly notes, phenomenal time (temporal experience) and phenomenal space (spatial experience) differ in this regard: there is no temporal analogue of our squinting and the like to phenomenally diminish the temporal analogue of our visual field. Nor do we know of any temporal analogue of medical conditions – such as hemianopsia

¹³ Thanks to an anonymous reviewer for pressing this point.

¹⁴ For possible senses in which James and his follower Stern can be interpreted to understand the notion of specious present, see Shardlow (2020). Thanks to an anonymous reviewer for making this point.

– that impair or reduce parts of our visual field of space. I am as helpless to phenomenally reduce the size of my temporal window of consciousness as I am to eliminate phenomenal time entirely from my consciousness. Admittedly it is an open question whether different specious presents might have different durations. Yet whenever a specious present is given, it must be given as a whole, and we cannot phenomenally reduce it to a shorter conscious interval. All we can do is to abstract from it to form a conception of a shorter conscious interval as its constituent.

One might also object that it is an open possibility that my introspective mode of presentation misrepresents the specious present as a unified whole: the present could be more fine-grained than it appears, with the shorter constituents simply given in phenomenal consciousness, so that the conception of any arbitrarily small part of it need not appeal to ideal abstraction. This objection is problematic for two reasons. First, it suggests that it is possible for phenomenal consciousness to appear other than what it really is like even upon conscientious introspection. However, it is questionable whether it makes sense to distinguish between the being and the appearing of phenomenal consciousness in this context. Most plausibly, what phenomenal consciousness is like simply is what it appears to be like upon conscientious introspection.

Second, in order for this objection to have appreciable force, the open possibility has to be a sufficiently serious one. Admittedly, the objector might remind us that there are credible cases where our introspective mode of presentation fails to represent all the significant but nuanced or fine-grained phenomenal features of experience. For instance, a non-expert in wine-tasting might not introspectively notice the rich layers of taste in her sip of a pinot noir. In cases of this

sort, the experience in question has nuanced or fine-grained phenomenal features that the introspective mode of presentation fails to represent. Citing such cases, the objector might contend that we should take seriously the open possibility that my specious present could in fact be more fine-grained than it appears.

However, the case of the specious present is different in kind from the case of wine-tasting. The key reason why we think of the case of wine-tasting as a credible case where the experience does indeed have phenomenal features that the introspective mode of presentation fails to represent is that this hypothesis is empirically confirmable. We have seen that, for instance, many students, who are non-experts in wine-tasting, having undergone the requisite training in a wine-tasting course, have come to be sensitive to the rich layers of taste of the pinot noir. Yet no empirical evidence of this sort exists in the case of the parallel hypothesis concerning the specious present. We know of no person or training or method such that she could come to represent the specious present as more fine-grained. The hypothesis that the specious present is in fact fine-grained in a way that escapes my conscientious introspection receives no confirmation from any empirical evidence of the sort that would make it a serious open possibility. The overall evidence favors the claim that specious present is a phenomenal unit whose parts can be conceived only through the whole. Therefore, most plausibly, the specious present is mereologically inverted.

V. Mereological Discrepancy between the Mental and the Physical

How does the mereological inversion of the specious present bear on the nature of the mind, and in particular, physicalism about the mind? The traditional debate concerning physicalism has been focused on the difficulty in explaining phenomenal consciousness in physicalistic terms. The main fight has been about the soundness of the knowledge argument, developed by Frank Jackson (1982) and (1986), and the conceivability argument, defended by David Chalmers (1996) and (2006). Both arguments exploit the explanatory gap between the physical and the phenomenal to the disadvantage of physicalism.

While I agree that the problem of phenomenal consciousness is highly important, I think the mereologically inverted structure of phenomenal time also poses a threat to physicalism in an interestingly different way. The difficulty stems from the fact that any physical particular is not mereologically inverted. Consider an arbitrary physical item, for instance, a neural network. Intuitively, the whole network ontologically depends on the constituent microphysical particles, which include numerous molecules, atoms, etc. Intuitively, the existence of the network depends on the existence of its constituents arranged in a certain way, in the sense that once these particles are there, and arranged this way, the neural network is there as a result. Therefore, most plausibly, no physical particular is mereologically inverted. However, something that is mereologically inverted cannot be purely constituted by or identical to something that is not mereologically inverted. Therefore,

NIC: The specious present cannot be purely constituted by or identical to any physical particular because of the mereological discrepancy between them.

I think NIC raises a serious challenge to physicalism. One way of formulating physicalism is to specify an identity relation between the mental and the physical. Thus, one might say that (a) every mental particular is identical to a physical particular, as in Donald Davidson's (1970) token-identity theory, or (b) every mental property is identical to a physical property, as in Jaegwon Kim's (1992) type-identity theory. The type-identity theory is stronger than the token-identity theory: the former entails the latter but not *vice versa*.

Another way of formulating physicalism appeals to the constitution relation. Thus, one might say that (c) every mental particular is purely constituted by a physical particular, or (d) every instantiated mental property is purely constituted by some physical property (or properties), as in Pereboom (2011), who holds that mental properties are high level compositional properties purely constituted by low level physical properties.

However, recall that by NIC there is at least one mental particular, i.e. the specious present, that cannot be identical to or purely constituted by any physical particular. Therefore, (a), (b), and (c) are in immediate trouble. What about (d)? Since the specious present is a particular that cannot be purely constituted by any physical particular, so the mental property of being specious-presently conscious cannot be purely constituted by the physical properties instantiated by any physical particular. Therefore, (d) is false. It seems that there is no plausible variant of physicalism appealing to the identity or constitution relation that can be true. The problem can be extended to bear upon variants of physicalism that appeal to the grounding relation (e.g. Shamik Dasgupta [2015]), or to the realization relation (e.g. Andrew Melnyk

[2003], Sydney Shoemaker [2007], Stoljar [2010], and Jessica Wilson [1999] and [2011]).

One might object that there are physical processes and systems that are indeed mereologically inverted. As the objection goes, for instance, my body's being in a certain orientation at a given moment can be an instantaneous time-slice of walking only if it is embedded within a temporal series of relevant bodily orientations. For another instance, assuming that functionalism is true, a certain firing of neurons constitutes a certain mental state only if that firing is appropriately embedded within a larger functional system.¹⁵

However, there is an obvious sense in which these processes are *not* mereologically inverted, and it is this sense that is relevant to my argument here. There is an evident sense in which walking is made possible by my sequentially arranging my body in the appropriate series of bodily orientations, and in this sense walking depends on the individual states that constitute it. For these individual states could each exist without their constituting this instance of walking, but this instance of walking could not exist without its being constituted by these individual states. Of course, it is trivially true that any such state cannot be an instantaneous time-slice of this instance of walking unless this state is what partially constitutes this instance of walking. However, this by no means suggests that the process of walking is mereologically inverted in the appropriate sense.

Moreover, there is an obvious sense in which a physical functional system, such as a car or a

¹⁵ Thanks to an anonymous reader of an earlier draft of this article for raising this objection.

brain, depends on its constituent parts. A car is manufactured by assembling its constituents on which its function depends: the tire does not depend on the car but the car depends on the tire. In a similar sense the function of a brain depends on the brain matter that composes it. This is generally true of physical processes and systems. However, the specious present is different because of its mereological inversion in the appropriate sense: none of its constituent durations could exist on its own in the way a tire or a small chunk of brain matter could.

In conclusion, there is a relevant sense in which there is a mereological discrepancy between the specious present and any physical particular, whether it is an object, a functional system, or a process. Physicalism about the mind runs into trouble because it has difficulty bridging the mereological gap between the mental and the physical.

Above I have assumed that what is physical is not mereologically inverted. Yet this assumption could be denied. Recently some, most notably Jonathan Schaffer (2010) and Schaffer and Ismael (2020), argue on the basis of quantum mechanics – e.g. the phenomenon of quantum entanglement – that the entire physical world is holistically grounded in the wave function. However, this view remains controversial. And even if physical holism is true, much work needs to be done to flesh out how specifically physical holism explains a mereologically inverted specious present. This would be an interesting topic for future research.

WORKS CITED

- Arstila, V. (2017) Temporal experience without the specious present, *Australasian Journal of Philosophy*, **96** (2), pp. 287-302.
- Broad, C. (1923) *Scientific Thought*, London: Routledge and Kegan Paul.
- Chalmers, D. (1996) *The Conscious Mind*, Oxford: Oxford University Press.
- Chalmers, D. (2006) Phenomenal concepts and the explanatory gap, in Alter, A. & Walter, S. (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: Oxford University Press.
- Chuard, P. (2011) Temporal experiences and their parts, *Philosophers Imprint*, **11**, pp. 1-28.
- Dainton, B. (2001) *Stream of Consciousness: Unity and Continuity in Conscious Experience*, London and New York: Routledge.
- Dainton, B. (2011) Time, passage, and immediate experience, in Callender, C. (ed.) *Oxford Handbook of Philosophy of Time*, Oxford: Oxford University Press.
- Dasgupta, S. (2015) The possibility of physicalism, *Journal of Philosophy*, **111** (9), pp. 557-592.
- Davidson, D. (1970/1980) Mental events, in *Essays on Actions and Events*, Oxford: Oxford University Press.
- Dennett, D. (1991) *Consciousness Explained*, London: Allen Lane.
- Einstein, A. (2005) Zur Elektrodynamik bewegter Körper, *Annalen der Physik*, **17**, pp. 891–921.
- Foster, J. (1982) *The Case for Idealism*, London: Routledge and Kegan Paul.
- Hershfield, H. E. & Maglio, S. J. (2020) When does the present end and the future begin? *Journal of Experimental Psychology: General*, **149** (4), pp. 701-718.
- Hoerl, C. (2013) A succession of feelings, in and of itself, is not a feeling of succession, *Mind* **122**, pp. 373-417.
- Husserl, E. (1893-1917/1991) Brough, J. B. (tr.). *On the Phenomenology of the Consciousness of Internal Time (1893-1917)*, Dordrecht: Kluwer Academic Publishers, 1991.

Ismael, J. (2011) Temporal experience, in Callender, C. (ed.) *Oxford Handbook of Philosophy of Time*, Oxford: Oxford University Press.

Jackson, F. (1982) Epiphenomenal qualia, *Philosophical Quarterly*, **32**, pp. 127-136.

Jackson, F. (1986) 'What Mary didn't know, *The Journal of Philosophy*, **83** (5), pp. 291-295.

James, W. (1890/1950) *Principles of Psychology*, Mineola: Dover Publications.

Kant, I. (1781/1787/1996) Pluhar, W. (tr.) *Critique of Pure Reason*, Indianapolis/Cambridge: Hackett.

Kim, J. (1992) Multiple realizability and the metaphysics of reduction, *Philosophy and Phenomenological Research*, **52**, pp. 1-26.

Koch, C. (2004) *The Quest for Consciousness*, Roberts: Colorado.

Le Poidevin, R. (2007) *Images of Time*, Oxford: Oxford University Press.

McKenna, C. (2020) Don't go chasing waterfalls: motion aftereffects and the dynamic snapshot theory of temporal experience, *Review of Philosophy and Psychology*, **12**, pp. 825-845.

Melnyk, A. (2003) *A Physicalist Manifesto: Thoroughly Modern Materialism*, Cambridge: Cambridge University Press.

Miller, K., & Holcombe, A. & Latham, A. J. (2020). Temporal phenomenology: phenomenological illusion versus cognitive error, *Synthese*, **197**, pp. 751-771.

Mundle, C. (1966) Augustine's pervasive error concerning time, *Philosophy*, **41** (156), pp. 165-168.

Pelczar, M. (2015) *Sensorama: A Phenomenalist Analysis of Spacetime and Its Contents*, Oxford: Oxford University Press.

Pereboom, D. (2011) *Consciousness and the Prospects of Physicalism*, Oxford: Oxford University Press.

Phillips, I. (2010) Perceiving temporal properties, *European Journal of Philosophy*, **18** (2), pp. 176-202.

Prosser, S. (2018) *Experiencing Time*, Oxford: Oxford University Press.

Rashbrook, O. (2012) Broad's accounts of temporal experience, *Journal for the History of Analytical Philosophy*, **1** (5): pp. 1-26.

- Russell, B. (1914) *Our Knowledge of the External World*, London: George Allen and Unwin.
- Schaffer, J. (2010) Monism: the priority of the whole, *Philosophical Review*, **119** (1), pp. 31-76.
- Schaffer, J. & Ismael, J. (2020) Quantum holism: nonseparability as common ground, *Synthese*, **197**, pp. 4131-4160.
- Shardlow, J. (2019) Minima sensibilia: against the dynamic snapshot model of temporal experience, *European Journal of Philosophy*, **27**, pp. 741-757.
- Shardlow, J. (2020) A tale of two Williams: James, Stern, and the specious present, *Philosophical Explorations*, **23** (2), pp. 79-94.
- Shoemaker, S. (2007) *Physical Realization*, Oxford: Oxford University Press.
- Soteriou, M. (2010) Perceiving event, *Philosophical Explorations*, **13** (3), pp. 223-241.
- Soteriou, M. (2013) *The Mind's Construction: The Ontology of Mind and Mental Action*. Oxford: Oxford University Press.
- Stoljar, D. (2010) *Physicalism*, New York: Routledge.
- Tye, M. (2003) *Consciousness and Persons*, Cambridge: MIT Press.
- Wilson, J. (1999) How superduper does a physicalist supervenience need to be? *Philosophical Quarterly*, **49**, pp. 33-52.
- Wilson, J. (2011) Non-reductive realization and the powers-based sub-set strategy, *Monist*, **94**, pp. 121-154.

CHAPTER III

Introspective Pessimism and the Reliability and Potency of Introspection:

Towards a Plausible Principle of Reliable Introspection

I. Introduction and Plan of the Chapter

As I am writing this article in my study, in the background plays the adagio movement ('Heiliger Dankgesang') from Beethoven's String Quartet No. 15. As I listen to the notes slowly unfolding their contemplative grandeur and transcendent tranquility, I find myself in the warm embrace of joyful serenity. Now I am enjoying a phenomenal experience that only I know what it is like to undergo. I know this in a way that is available only to me and not to others. As I turn my attention inward to examine my present phenomenal experience – as I try to introspect, in a sense to be specified later – can I be mistaken about what it is like to undergo it? And do I know all or only some aspects of what it is like to undergo it? More generally, for any ongoing phenomenal experience of mine, (a) how reliably do I know by introspection what it is like to undergo it? And (b) how much do I know by introspection what it is like to undergo it?

Set question (b) aside for the moment. To question (a), there was a long and honorable tradition in the history of philosophy that gave a definitively negative answer: I cannot be mistaken and cannot stand corrected by others. On this view, introspection is maximally reliable – i.e. infallible – and incorrigible – i.e. not subject to appropriate third personal correction – in telling me what my ongoing phenomenal experience is like. The tradition has an honorable intellectual

pedigree that includes, to name a few illustrious thinkers, Augustine of Hippo (420/2003) of the late antiquity, Rene Descartes (1641/2017) of the early modern period, and Edmund Husserl (1913/1982) and A. J. Ayer (1936/1946) of the modern period.

However, this tradition came under attack during the first half of the 20th century with the rise of behaviorist psychology. The psychologists in the lab were astonished and scandalized by the fact that the introspective reports of their subjects contradicted with one another on various occasions. As a result, they could not help viewing introspection and its reliability with severe suspicion. The taint of its ill repute remained even when behaviorism in psychology and philosophy of mind declined (Hohwy [2011]). In the last few decades, introspection found its several of its most vigorous critics in philosophy, among whom are Daniel Dennett (1991) and (2002), and more recently, Eric Schwitzgebel (2008) and (2011). However, introspection was not without its defenders, featuring, just to name a few representatives, David Chalmers (2003), Terry Horgan and Uriah Kriegel (2007), Jakob Hohwy (2011), Declan Smithies (2013), Michael Pelczar (2014), Timothy Bayne and Maja Spener (2015), and Uriah Kriegel (2015).

In this article I examine three attitudes towards how well and how much introspection reveals about what our phenomenal experience is like: pessimism, optimism, and moderatism. These attitudes are not meant to be exhaustive. However, I think they are worth critical discussion due to their representativeness. I begin by introducing two key terms by which I will structure my subsequent discussion: (a) the reliability and (b) the potency of introspection (Kriegel [2015]). Specifying the distinction is the task of the following section, and here suffice it to say

that the more (a) reliable introspection is, the more trustworthy its verdicts are about our phenomenal consciousness; and the more (b) potent it is, the more it reveals about our phenomenal consciousness. Having set the stage for my discussion with these notions, I go on to examine a pessimist, an optimist, and a moderate outlook on the reliability and the potency of introspection. The key lesson will emerge that a full appreciation of this distinction has important revelations. For one, many of the pessimistic concerns with introspection threaten not so much the reliability as the potency of introspection. For another, once the two notions are disentangled, the reliability of conscientious introspection emerges as eminently defensible.

Several clarifications are in order. By introspection, I mean our peculiar first personal access to our phenomenal conscious states. Here ‘peculiar’ is intended in the sense of Alex Byrne (2018): the access is peculiar insofar as it is available only to the subject but not to others. As Byrne observes, peculiarity is to be distinguished from privilege. Having a peculiar access does not entail that the access is epistemically privileged with regard to either its reliability or its potency. What the peculiar access consists in is difficult to articulate in positive terms. However, for the current purpose, let the example in the beginning paragraph suffice: I have epistemic access to my ongoing experience in a peculiar way that you do not. This negative strategy of characterizing introspection is adopted by Byrne (2018) and Silins (2020) among others.

I think of introspection as a judgment generating process that takes as its input the phenomenal consciousness and yields a judgment on the basis of this input in a peculiar first personal way.

I leave open the question of whether such a functionally defined process could be multiply

realized by diverse mechanisms as urged by Schwitzgebel (2012). Moreover, on my usage, a *purely* introspective belief that an experience *e* instantiates a phenomenal property *Q* is a belief immediately based on *e* with the minimal conceptual resources sufficient to express the belief ‘*e* is *Q*’. Introspective purity comes in degrees. An introspective belief is *impure* to the extent that its relation to experience is mediated and influenced by background beliefs.

Finally, the scope of introspection under discussion is intentionally narrow. I will focus only on our introspective access to our phenomenal consciousness with a view to questions (a) and (b). I will not discuss introspection in the context of the self-ascription of propositional attitudes such as beliefs, which are the primary concerns for Gareth Evans (1982), Tyler Burge (1996), Richard Moran (2001), and Byrne (2018). Also I will not examine introspection in the context of our self-ascription of desires, motives for actions, psychological dispositions, mental habits, personality traits, etc. Finally, I will not be concerned with the subconscious, steering clear of the muddy and dangerous Freudian waters.

II. Introspective Reliability and Potency

Let us begin by reminding ourselves of Uriah Kriegel’s (2015) very useful distinction between the reliability and the potency of introspection.

Introspective Reliability: Given that the subject *S* introspects an experience *e* of hers as instantiating a phenomenal property *Q*, how likely it is that *e* really instantiates *Q*.

By contrast:

Introspective Potency: Given that the subject S has an experience e that instantiates a phenomenal property Q, how likely it is that S introspects e as instantiating Q.

Let 'an experience e that instantiates a phenomenal property Q' be abbreviated as 'a Q-phenomenology'. Thus:

The more reliable introspection is, the more likely it is that I do have a Q-phenomenology if I introspect one.

And:

The more potent introspection is, the more likely it is that I do introspect a Q-phenomenology if I do have one.

As mentioned earlier, many philosophers in the history have claimed maximal reliability for introspection.

Maximal Reliability: If the subject S introspects a Q-phenomenology, then she really has a Q-phenomenology.

While it used to be a popular position to claim maximal reliability for introspection, few have claimed maximal potency for it.

Maximal Potency: If the subject S is having a Q-phenomenology, then S introspects having it.

It is not difficult to see why MP has few adherents now. The key reason is that MP implies two problematic claims. MP implies, quite implausibly, that first, I carry out acts of introspection throughout my entire conscious mental life, and second, my acts of introspection always reveal the phenomenology of my conscious mental life in its entirety. Since every act of introspection

most plausibly involves introspective attention, and since there are numerous occasions on which I am far from being introspectively attentive to what is going on phenomenologically within my stream of consciousness, it is implausible that I keep introspecting throughout my entire conscious mental life (*contra* the first claim). Moreover, since attention of any kind most plausibly has a limited scope, so that I am usually unable to attend to everything available for my attentive scrutiny, it is implausible that my conscious mental life always reveals its phenomenal character in every aspect and in its entirety (*contra* the second claim).

Several clarifications are in order. First, what is at stake is how good introspection alone is with regard to our conscious mental life. The questions with which we are concerned are how reliable introspection is (its reliability), and how much introspection reveals of our conscious mental life (its potency), apart from the perceptual, doxastic, or any other influences from any non-introspective capacities and their operations. This point is often noted in the contemporary literature. For instance, Bayne and Spener (2010) cites what they call the ‘influence from background belief and expectation’ as an important source of introspective disagreements. Strategies of the same spirit are developed by many others such as Siewert (2007) and Smithies (2013). In my terminology, these influences *taint* introspective beliefs by making them *impure*.

Second, what is at stake is how good introspection is with regard to ongoing experience. It is irrelevant to our evaluation of the reliability and the potency of introspection how reliable and potent introspection is with regard to past experiences. It is a familiar point, going back at least to Descartes (1641/2017) and noted by contemporary commentators such as Barry Stroud

(1984), that introspection grants us privileged epistemic access to our ongoing experience. In fact, I remain skeptical that at any given time I can strictly speaking introspect past experiences.

These observations have important ramifications: they remind us of the key constraints on what kind of problems with certain first personal reports bear on the reliability and the potency of introspection alone. Let us look at two kinds of problems, which, despite initial appearances, do not really bear on the reliability or the potency of introspection.

The first kind of problem consists in the difficulty in reliably reporting the coloration of our dreams. Are our dreams colorless, or only black and white, or multiply colored? Schwitzgebel (2011) argues that it is difficult to tell. He notes that first personal reports vary quite widely both diachronically – i.e. throughout different historical periods – and i.e. synchronically – among contemporary subjects. He then argues against the possibility that the coloration of their dreams varies just as much. If so, then at least a significant body of reporting subjects must be wrong. Moreover, it is difficult to ascertain which the correct reports are.

Here I will not dispute Schwitzgebel's empirical case for his agnosticism. Instead, let me point out that the difficulty in telling how our dreams are colored, while it could be seen as evidence for our lack of knowledge of a significant aspect of our mental life, does not in fact bear on the reliability and the potency of introspection. First, at the time of these reports, since the subjects are now awake, dreams are no longer their ongoing experiences. Second, the reporting subjects rely heavily on memory, of whose reliability we have independent reasons to remain skeptical.

Third, it is likely that the reports are contaminated by *doxastic* influences. Schwitzgebel himself notes that there are telling correlations between the coloration of the media to which the subjects are frequently exposed and the coloration that they tend to report their dreams to have. He thinks it is unlikely that the coloration of dreams varies phenomenally across these subjects. If so, I think the best explanation is that the subjects fall prey to the non-introspective *doxastic* influences: their reports likely reflect their background beliefs which are heavily influenced by the coloration of their familiar media. With the contamination of such doxastic influences, the introspective beliefs whose reliability is in question are impure at best. Given their impurity, they are no evidence for the unreliability of introspection.

Schwitzgebel does not claim that such difficulties suffice to undermine the reliability and the potency of introspection. He merely says, correctly I think, that it falls into the same general category of cases that suggest that we are significantly more ignorant about our mental life than we think. However, I think the cases surveyed so far offer no reason to blame this ignorance on the defects of introspection.

The second kind of problem that merely appears to bear on the reliability and the potency of introspection consists in the dubious reliability of the first personal reports about whether there are what I will call ‘peripheral phenomenologies’, i.e. phenomenologies putatively present but beyond the sphere of attention. Here is Schwitzgebel’s list of questions:

Do you have constant tactile experience of your feet in your shoes? Constant auditory

experience of the hum of traffic in the background? Constant visual experience of the frames of your eyeglasses? Or, when you aren't attending to such matters, do they drop out of consciousness, so that they are in no way part of your stream of experience, your phenomenology? Is consciousness abundant, the stream of experience bristling with phenomenology in a wide variety of modalities simultaneously (visual, auditory, tactile, olfactory, imagistic, proprioceptive, emotional), or is it sparse, limited to one or a few things at a time? (Schwitzgebel 2011: 91).

Schwitzgebel argues that the reports obtained *via* what seems to be most promising method of introspection are not reliable, and we are ignorant of whether peripheral phenomenologies exist.

The introspective method is as follows:

We might try giving participants beepers to wear during their normal daily activities – beepers that sound only at long intervals, when the participants are likely to be immersed in other things – and instructing the participants to reflect, each time the beeper sounds, on what their experience was immediately before the beep, when (in most cases, presumably) they won't have been thinking about the sparseness or abundance of experience, or about their feet, or about the traffic in the background (Schwitzgebel 2011: 99).

Schwitzgebel argues that in the end, it is doubtful that the reports are reliable, and I concur with his judgment.

However, my point here is different: even if we grant this, we are not warranted to further claim that the unreliability of these reports is somehow evidence for the unreliability or the potency

of introspection. The subjects are asked to reflect on what the experience is like immediately before the beep. It is not the on-going experience on which the subjects are asked to reflect. At the moment of the beep, what the subjects end up reflecting on is what they remember to have experienced. Moreover, it is unclear that the reports are not contaminated by the subjects' prior beliefs about whether, and which kinds of, peripheral phenomenologies there are. Those who believe that conscious experience is rich would tend to report the existence of peripheral phenomenologies, while those who do not would not. Absent further arguments, it is hard to see how the unreliability of these reports bears on the reliability of introspection. Rather they seem to bear more on the reliability of memory and prior beliefs.

Before moving on to consider Schwitzgebel's own introspective pessimism, let me consider a line of argument against the reliability of introspection. The argument is not Schwitzgebel's own, but it owes its inspiration to a remark of his. At one point he argues that when we form a visual image using our imagination, it is in fact very difficult to know, on the basis of introspection, what our visual imagery experience is like beyond a bare level of sketchiness.

Close your eyes and form a visual image. [...] Imagine [...] your breakfast table as you sat down to it this morning. Or imagine the front of your house as viewed from the street. Assuming that you can form such imagery (some people say they can't), consider this: How well do you know, right now, that imagery experience? You know, I assume, that you have an image, and you know some aspects of its content – that it is your house, say, from a particular point of view. But that really isn't to say very much about your imagery experience.

Consider these further questions: How much of the scene can you vividly visualize at once? Can you keep the image of the chimney vividly in mind at the same time that you vividly imagine your front door, or does the image of the chimney fade as you begin to think about the door? How much detail does your image have? How stable is it? If you can't visually imagine the entire front of your house in rich detail all at once, what happens to the aspects of the image that are relatively less detailed? If the chimney is still experienced as part of your imagery when your image making energies are focused on the front door, how exactly is it experienced? Does it have determinate shape, determinate color? [...] If there is indeterminacy of color, how is that indeterminacy experienced? As gray? [...] (Schwitzgebel 2011: 36).

And the list of questions goes on. Schwitzgebel points out that many of us find it difficult to answer these questions with any confidence. This difficulty undermines the general optimism in our knowledge of our own ongoing conscious experience. I think Schwitzgebel is right here.

Although Schwitzgebel does not explicitly frame this as an attack on introspection, one might be tempted to do so by arguing that our proneness to errors about ongoing visual imagery seems to be evidence against the reliability of introspection. However, I think this argument faces a serious difficulty. The key issue here is whether the visual imagery itself is determinate or not. Drawing on George Berkeley's (1710/1734/1982) famous imagistic conception of ideas against John Locke's (1690/1979) abstractionism, a Berkeleyan on visual imagery might say that whenever I visually imagine – call to mind an idea of – a chimney, its visual imagery – its idea – is concrete and determinate down to a certain level of detail: the chimney has a concrete and

determinate contour, features concrete and determinate colors distributed in a concrete and determinate pattern, etc. However, this Berkeleyan conception could be plausibly resisted. The visual imagery might be concrete and determinate to a high level in certain aspects while not so in others. If so, there might simply be no matter of fact, for instance, of which precise and determinate angles the chimney really has.

Acknowledging this possibility forces a dilemma upon the critic of introspective reliability. Suppose that the phenomenology is itself non-concrete and indeterminate (and this is the first horn of the dilemma). Even so, it is unclear that introspection really fails here, for it delivers a result no less non-concrete and indeterminate than the actual phenomenology. To be sure, the chimney is a metaphysically determinate object. However, there might not be a phenomenally determinate visual imagery of it to introspect with any high level of phenomenal concreteness and determinacy. Schwitzgebel might be correct that I could easily be mistaken about what my experience is like. However, my proneness to mistakes likely derives from my misplaced effort to report, with greater concreteness and determinacy than there really are in my phenomenology, what the non-concrete and indeterminate aspects of my visual imagery are like. This misplaced effort is likely the result of a misleading doxastic influence that confuses an indeterminate phenomenology with determinate objects themselves.

Now suppose, for the sake of argument, that the phenomenology is itself determinate (and this is the second horn of the dilemma). The problem here would be that introspection is too weak to deliver any indeterminate result. It would be a failure of the following sort: if I have a Q-

phenomenology, I fail to (concretely and determinately) introspect one. It would not be a failure of the following sort: if I (concretely and determinately) introspect a Q-phenomenology, I do not in fact have one. It would be the potency, rather than the reliability, of introspection that the putative evidence in question counts against. However, I see no reason to grant that the phenomenology is concrete and determinate to a high level in the first place. Therefore, neither the reliability nor the potency of introspection is in serious threat here.

Another possible response, suggested by Hohwy (2011) and elaborated on by Bayne (2015), is that the phenomenology itself is variable in the sense of constantly shifting, and the subjects introspective it as so, and hence reliably. As Bayne (2015) suggests:

Perhaps we are not sure how best to describe the phenomenology of imagery because it is so variable. Imagery experiences cannot be pinned down, but are constantly shifting in response to our own imagistic activity. Precisely how much of the scene we vividly visualize ‘all at once’ depends on the goals that constrain the act of visualization. And, as Hohwy suggests, when we have no such goals our imagery may end up ‘freewheeling’, such that we move from one state to another. Hohwy grounds his analysis in a predictive-coding account of cognition, but his fundamental point is independent of that theoretical framework and should be fairly uncontentious: imagery surely is more labile than perceptual experience or bodily sensation. No wonder, then, that its phenomenal structure is that much more difficult to articulate (Bayne 2015: 10).

Despite my sympathy, I do not think this move is either necessary or convincing. It is not clear

to me that the phenomenology itself is concrete and determinate at all, and it seems implausible to posit a shifting series of determinate images in rapid succession. I doubt that our minds, when left on their own unassisted by any visual media, have so strong an imaginative power as to produce even one fully concrete and determinate Berkeleyan visual image.

III. Introspective Pessimism: Eric Schwitzgebel

Above we have seen that Schwitzgebel's discussion of introspecting the phenomenology of visual imagery does not easily lend itself to an argument that undermines the reliability or the potency of introspection. In none of his remarks considered so far does Schwitzgebel mount an explicit attack against introspection. Indeed, I have been arguing in length that none can easily be turned into one. Now, switching gears, I proceed to consider what Schwitzgebel explicitly frames as critiques of introspection.

Schwitzgebel raises two kinds of cases. The first kind of cases are examples of the type that introspection fails to tell us about certain properties of our experience. Here is a typical passage from Schwitzgebel's explicit critique of naïve introspection:

You've had emotional experiences, and you've thought about them, reflected on how they feel as they've been ongoing or in the cooling moments as they fade. If such experiences are introspectible, and if introspection is the diamond clockwork often supposed, then you have some insight. So tell me: Are emotional states like joy, anger, and fear always felt phenomenally – that is, as part of one's stream of conscious

experience – or only sometimes? Is their phenomenology, their experiential character, always more or less the same, or does it differ widely from case to case? For example, is joy sometimes in the head, sometimes more visceral, sometimes a thrill, sometimes an expansiveness – or, instead, does joy have a single, consistent core, a distinctive, identifiable, unique experiential character? (Schwitzgebel 2011: 120).

In reply, let me begin with a preliminary point: Schwitzgebel does not sharply distinguish between the reliability and the potency of introspection. Even granting for the sake of argument that it were somehow the fault of introspection that we cannot answer these questions, it should be clear that it is not really the reliability, but rather the potency, of introspection that is under threat. The difficulties or failures that Schwitzgebel mentions above would merely show that introspection is too impotent to deliver any unambiguous results.

However, a more substantial point is in order: despite initial appearance, these difficulties or failures do not really bear on either the reliability or the potency of introspection. The reason is that these issues are not questions of what our conscious experience is like phenomenally: they are not even phenomenal questions that introspection is purportedly in the business of answering. Rather, they are further theoretical questions about how the instances of given phenomenal properties relate to one another. Interesting though they are, they have no obvious bearing on either the reliability or the potency of introspection.

An analogy might be helpful. I take it to be obvious that neither the reliability nor the potency

of *perception* is undermined by our failure to tell whether there is a single, unique color-character across all perceived colors or what that common color-character is. Similarly, neither the reliability nor the potency of *introspection* should be undermined by our failure to tell whether there is a single, unique phenomenal character across all phenomenal states of the same type or what that common phenomenal character is. Therefore, the first kind of cases Schwitzgebel raises are irrelevant for our purposes.

In his reply to Schwitzgebel, Pelczar (2014) seems to be hinting at exactly this point, but I think it is entangled with other points that are less than convincing. Here is Pelczar's reply:

Introspection alone cannot possibly tell me whether all my states of joy have a phenomenal character in common, for the simple reason that I cannot introspect all my joyful states together. To answer the questions that Schwitzgebel asks here, I must, at a minimum, draw on the resources of memory (and probably other cognitive capacities as well, such as inference and analysis). If I am unsure how to answer these questions [...], the most that this shows is that a combination of introspection and other cognitive functions is fallible. In particular, the claim that judgments about whole kinds of phenomenal experience are fallible is consistent with introspection being entirely infallible [...] (Pelczar 2014: 20-21).

Pelczar is quite right that it requires much more than introspection to answer Schwitzgebel's questions. However, even if I could introspect all my states of joy together, I might still find it difficult to say whether there is a common phenomenal character, and more importantly, I claim that this difficulty still would not bear on the reliability of introspection. I find the task difficult

because any conceptual generalization that aims at distilling a common defining feature of a class of entities – even that of a natural kind – is very demanding. However, this difficulty has very little to do with the reliability of the immediate mode of presentation through which these entities are presented. Again, compare the case of perception. Even if I could perceive all colors together, I might still find it difficult to tell whether there is a common color-character and what that common color-character is, and I take it to be obvious that this difficulty has no bearing on the reliability of visual perception. The same holds for introspection.

However, the second kind of cases that Schwitzgebel raises are indeed relevant to the reliability of what he calls ‘naïve introspection’. As I read through the previous sentence, *bona fide* I conscientiously introspect my visual experience. However, what I do on this occasion does not embody any sophisticated guidelines or reflect any disciplined practice. It is in this sense that our introspection is naïve.

Like most people casually reflecting on their conscious visual experience, I naively believe that my visual experience is stable, and the area of clarity of my visual field is large. However, Schwitzgebel contends that we are mistaken, and our naïve introspection is far from reliable. His argument is that through a set of exercises with relatively sophisticated guidelines, we will eventually come to discover that, to our surprise, our visual experience is not what we think it is like at all. Focusing on the case of peripheral visual phenomenology discussed in Dennett (1991) and (2002) and Blackmore (2002), Schwitzgebel says:

Fixate on some point in the distance, holding your eyes steady while you reflect on

your visual experience outside the narrow fovea. Better, direct your introspective energies away from the fovea while your eyes continue to move around (or “saccade”) normally. This may require some practice. You could begin by keeping one part of your visual field steadily in mind while you allow your eyes to foveate anywhere but there. Take a book in your hands and let your eyes saccade around its cover while you think about your visual experience in the regions away from the precise points of foveation (Schwitzgebel 2011: 126).

The result is that:

Most of the people I have spoken to who attempt these exercises eventually conclude to their surprise that their experience of clarity decreases substantially even a few degrees from the center. Through more careful and thoughtful introspection, they seem to discover – I think they really do discover – that visual experience does not consist of a broad, stable field, flush with precise detail, hazy only at the borders. They discover that, instead, the center of clarity is tiny, shifting rapidly around a rather indistinct background. Most of my interlocutors confess to error in having originally thought otherwise (*ibid.*).

Schwitzgebel concludes:

If I am right about this, then most naïve introspectors are badly mistaken about their visual phenomenology when they first reflect on it. Even though they may be patiently considering their experience as it occurs, they will tend to go wrong unless they are warned and coached against a certain sort of error. And the error they make is not a subtle one; the two conceptions of visual experience differ vastly. If naïve

introspectors are as wrong as many later confess themselves to be, they are wrong about an absolutely fundamental and pervasive aspect of their sensory consciousness (ibid.).

Earlier I emphasized that Schwitzgebel's target is *naïve* introspection. This qualification matters. What Schwitzgebel has succeeded in establishing, I think, is that a peculiar species of introspection is unreliable in a significant case. This rather modest claim is far from a whole scale rejection of introspection of any kind as unreliable. In fact, it is worth stressing that much of the force that Schwitzgebel's argument has against naïve introspection in this type of cases in fact depends on the assumed reliability of more sophisticated introspection, as noted in the contemporary literature, for instance in Hohwy (2001). It is the fact that this more sophisticated introspection yields a different result from naïve introspection that is supposed to offer us the key reason for believing that what our naïve introspection reveals of the stability and the area of clarity of our visual experience is in fact far from veridical. Many defenders of introspection, such as Charles Siewert (2007) and Pelczar (2014), have indeed tried to specify optimal conditions in which introspection remains reliable.

Having presented the argument against the reliability of naïve introspection, Schwitzgebel adds that he is willing to doubt it. He acknowledges that he remains open to the possibility that the area of clarity in our visual field might be as large as what naïve introspection reveals to us, and that he and many others might have erred in their critique of naïve introspection. However, he hastens to add that this concession presents no less a problem for introspection:

I am perfectly willing to doubt myself, though. Maybe I'm wrong and visual experience is broadly crisp and stable. But if so, I'm not the only person who is wrong about this. So also are most of my interlocutors (whom I hope I haven't browbeaten too badly) and probably a good number of philosophers and psychologists. We – I, my friends, and co-believers – have erred through some theory or preconception, perhaps, some blindness, stupidity, oversight, or suggestibility. Okay, let's assume that. I need only, now, turn my argument on its head. We tried to get it right. We reflected, sincerely, conscientiously, in good faith, at a leisurely pace, in calm circumstances, without external compulsion, and we got it wrong. Introspection failed us. Since what I am trying to show is the aptitude of introspection to lead to just such errors, that result would only further my overall thesis. Like other skeptical arguments that turn on our capacity for disagreement, it can triumph in partial defeat (Schwitzgebel 2011: 127).

The key idea here is that even if Schwitzgebel were wrong in his critique of the reliability of naïve introspection, introspection would still fail us with its lack of reliability.

In reply I would like to make two points. First, it remains true that the force of Schwitzgebel's earlier argument against the reliability of naïve introspection depends on the assumed reliability of more sophisticated introspection. Thus, any doubt of the reliability of the latter should entail a decrease in doubt of the reliability of the former. Schwitzgebel cannot have his arguments to undermine both kinds of introspection at once.

Second, absent any further arguments, I am unconvinced that this skepticism of the reliability of more sophisticated introspection should be taken seriously. Admittedly, we must not be dogmatic. However, I do not see any reason to doubt the reliability of the more sophisticated introspection. As Schwitzgebel acknowledges, not only is introspection of this sort carried out with great care, but people who have followed it through also come to the agreement that the area of clarity in our visual field is small and unstable; and finally, such results are duplicated on numerous occasions. All these points support the reliability of more sophisticated exercise.

Perhaps a sympathizer of Schwitzgebel might complain that even sophisticated introspection does not tell us what the optimal conditions are and whether they obtain at the time of its operation. Introspection does not present its results with labels indicating whether the conditions at the time of the results are optimal or not. Such limitations partially explain why we are so prone to errors when we introspect. However, to my mind this complaint asks too much of introspection. An indication of introspective conditions goes beyond the question of what our phenomenal experience is like: it is not a phenomenal question that introspection is in the business of answering. Moreover, this complaint does not tell against the reliability of sophisticated introspection. At most it tells against – with questionable force – its potency.

Let me briefly comment on what seems to me to be a misunderstanding of Schwitzgebel's view.

Pelczar (2014) reads Schwitzgebel as arguing along the following schema.

ID1. There are people who have experiences that are identical in a given phenomenal respect, but who disagree about the phenomenal character of their experiences in this

respect, even upon ideal introspection.

ID2. If ID1 is true, then some people mistake the phenomenal characters of some of their experiences, even upon ideal introspection.

ID3. So, ideal introspection is fallible (Pelczar 2014: 22).

I do not think this schema accurately captures how Schwitzgebel argues. Rather, as I read him, his claims are, first, that naïve introspection is unreliable, and second, that if his case against the reliability of naïve introspection failed, then it would still be introspection – i.e. of the more sophisticated sort – that failed. The interpersonal disagreement that Schwitzgebel exploits is really the one between the results delivered by naïve introspection and those delivered by more sophisticated introspection. To my knowledge, he never claims that people disagree about what their visual experience is like upon ideal introspection.

While I agree with Schwitzgebel that naïve introspection is sometimes unreliable, I do not think that we have any good reason to embrace his general pessimism about introspection. In fact, we have good reasons to trust the reliability of sophisticated introspection in optimal conditions.

IV. Introspective Optimism: Michael Pelczar

Unlike Schwitzgebel, Pelczar is an optimist concerning introspection. He professes himself a defender of the reliability of introspection appropriately exercised. His core position is that under ideal circumstances, non-confounding qualia are available for introspection. What does he mean by ‘ideal circumstances’ and ‘non-confounding *qualia*’? A necessary condition for a

subject to be under ideal circumstances relative to a quale is that she is cognitively unimpaired.

According to his usage, cognitive impairment

includes things like mental malfunction (due to fatigue, intoxication, etc.), inattention (due to laziness or carelessness), and psychological disturbance or disruption (due to being startled or frightened, for example) (Pelczar 2014: 13).

To be cognitively unimpaired is to be exempt from such mental malfunction, inattention, and psychological disturbance or disruption. Pelczar goes on to characterize what it is to introspect an experience *e* under ideal circumstances relative to a *quale* *Q*:

Suppose that in this unimpaired state, I introspect an experience, *e*, with a view to determining whether it has a certain phenomenal property, *Q*, and having taken all reasonable measures to ensure that my determination is correct. Then I am, as I shall say, introspecting *e* under ideal circumstances, relative to *Q* (ibid.).

Pelczar claims that it is the non-confounding *qualia* that are available to introspection under ideal circumstances. A *quale* is non-confounding just in case it is not a confounding one:

Q is a confounding quale just in case it is impossible to introspect any experience under ideal circumstances, relative to *Q* (ibid.).

And:

The most extreme example of a confounding quale is what we might call ‘phenomenal obliviousness’. This is a phenomenal property by virtue of instantiating which an experience is, among other things, an experience as of some state of affairs obtaining in the absence of any introspection (ibid.).

Phenomenal obliviousness is in principle unavailable to introspection under any circumstances.

With the key terms clarified, we are ready to understand Pelczar's Principle of Introspection (PI). I will label his formulation as PI-P.

PI-P: Every experience that I have, *e*, is such that for each of its non-confounding qualia, *Q*, (a) if *e* has *Q* and I introspect *e* while cognitively unimpaired, and with a view to determining whether *e* has *Q*, and having taken all reasonable measures to ensure that my determination is correct, then I introspect that *e* has *Q*, and, (b) if *e* does not have *Q*, and I introspect *e* while cognitively unimpaired, and with a view to determining whether *e* has *Q*, and having taken all reasonable measures to ensure that my determination is correct, then I introspect that *e* doesn't have *Q* (Pelczar 2014: 12).

Loosely speaking, PI claims that non-confounding *qualia* are available to introspection under ideal circumstances.

Despite Pelczar's claim that PI-P concerns reliability of introspection, it is in fact its potency that PI-P really concerns. What PI-P claims is not that under ideal circumstances, with the appropriate qualifications, if I introspect having a *Q*-phenomenology, I do in fact have one. What it claims instead is that under ideal circumstance, if I do in fact have a *Q*-phenomenology, then I introspect having one. It is the potency, rather than the reliability, of introspection that PI-P really concerns.

The reliability of introspection should not be confused with its potency. Neither of them entails the other. Even if PI-P, which asserts the potency of introspection, were true, it would still not

give any reason for believing that introspection is reliable. PI-P provides no reason for ruling out the possible scenario that my introspection errs frequently enough to count as unreliable: it could be the case that all non-confounding qualia of mine are indeed revealed, and yet these veridical results comprise merely a rather small fraction of the total results, of which many others are non-veridical. In this case, introspection would indeed reveal all the non-confounding *qualia* under ideal circumstances and is therefore highly potent, but it remains true that it would err on most occasions and is therefore highly unreliable. High potency is compatible with high unreliability.

This point matters, because there seems to be a problematic gap between what Pelczar thinks PI supports and what PI actually supports. As I read him, he argues for what is in effect the potency of introspection, and then proceeds argue in the following chapters of his book as if he has established what is in effect its reliability. Pelczar makes clear that he intends to show that introspection is reliable with due qualifications so that he can base many of his claims about what our conscious experience is like on introspection as it is appropriately exercised.

I base many of the claims that I make in this book on introspection. For example, I claim that introspection frequently reveals to me experiences as of things changing, and I say that this gives me every reason to think that I actually have such experiences. I also claim that I never introspect an experience as of a ten dimensional solid, or as of a million-sided polygon, or as of one sound lasting a nanosecond longer than another, and I say that this gives me every reason to think that I never have an experience as of a ten-dimensional solid, or a million-sided polygon, or one sound

outlasting another by one nanosecond (Pelczar 2014: 10).

Pelczar's basic position is that introspection is reliable when properly exercised. He proceeds to argue in this fashion: if I introspect that *e* is *Q* with due care, then *e* is indeed *Q*. Inferences of this sort are not the sort warranted by PI-P. By itself PI-P provides no support for the intended claim that the introspection-based claims of what our conscious experience is like are reliable or even more likely true. It is unclear what PI-P has to do with his professed intentions. It is true that Kriegel's work postdates Pelczar's, who might not be working with Kriegel's terminology. However, the philosophical problem for Pelczar remains as real as it is, even if it is fleshed out in Kriegel's terminology.

Perhaps Pelczar's intention is to defend a conjunction in favor of both the reliability and the potency of introspection, of which only the first half has been made explicit. The first conjunct, in favor of the potency of introspection under ideal circumstances, has been spelled out in PI-P. The second conjunct, which Pelczar most likely intends to defend but unfortunately does not include in his explicit formulation, is most likely a claim to the effect that introspection is reliable insofar as it does not err with regard to whether certain non-confounding *qualia* are present or not under ideal circumstances. I propose the following formulation as a friendly amendment to Pelczar's PI, using his terminology.

PI: PI is the conjunction of PI-P and PI-R articulated as follows:

PI-R: Every experience that I have, *e*, is such that for each of its non-confounding *qualia*, *Q*, (a) if I introspect that *e* has *Q*, and I do so while cognitively unimpaired, and with a view to determining whether *e* has *Q*, and having taken all reasonable

measures to ensure that my determination is correct, then e has Q, and, (b) if I introspect that e does not have Q, and I do so while cognitively unimpaired, and with a view to determining whether e has Q, and having taken all reasonable measures to ensure that my determination is correct, then e does not have Q.

Unlike PI-P, PI-R would provide the right kind of support for Pelczar's claim that introspection is reliable under ideal circumstances and thereby underwrite the key claims he draws on the basis of introspection. This suggests that PI-R is probably the correct amendment to his position.

Not a pessimist about introspection, I find PI-R very plausible. The opponent might complain that the clause of reasonable measures involves an ambiguity. On one reading, it requires that:

R1: The subject should undertake all the measures necessary for a reliable exercise of introspection.

However, the opponent might argue, R1 has two problems. The first problem is that of vacuity: it is trivially true that, according to PI-R, if the cognitively unimpaired subject has undertaken all measures necessary for a reliable exercise of introspection, her exercise of introspection will turn out to be reliable. The second problem is that of application: how can we come to know that we have taken all the measures necessary for a reliable exercise of introspection? If we cannot know this, then it seems that we are hardly ever in a good epistemic position to apply PI-R to specific cases so as to underwrite any interesting or substantial claims drawn on the basis of introspection. However, R1 could be easily amended as follows:

R2: The subject should undertake all the measures (a) which are known to be necessary for a reliable exercise of introspection, and (b) which she is capable of

carrying out.¹⁶

This reading avoids the problem that the first reading faces. Construed under R2, PI-R is very plausible even if not infallible.

Let me apply PI-R construed under R2 to some cases putatively problematic for introspection. Derk Pereboom (2011: 14) raises two interesting cases to support his thesis that it is an open possibility that introspection misrepresents our phenomenal experience as having phenomenal properties that it does not actually have. The first case, reported by Roger Albritton and cited by Christopher Hill (1991) and Pereboom (2011),

involves a college student who is being initiated into a fraternity. He is shown a razor, and is then blindfolded and told that the razor will be drawn across his throat. When he feels a sensation he cries out: he believes for a split second that he is in pain. However, after contemplating the sensation for a moment, he comes to feel that it is actually an experience of some other kind. It is, he decides, a sensation of cold. And this belief is confirmed when, a bit later, the blindfold is removed and he is shown that his throat is in contact with an icicle rather than a razor (Hill 1991: 128-129; cited in Pereboom 2011: 22).

Pereboom comments that:

one possibility is that in his introspective awareness, the fraternity pledge at first misrepresents the qualitative features of the sensation of cold he actually has as qualitative features of pain, and later it becomes clear to him that he has

¹⁶ Thanks to Derk Pereboom for the helpful suggestion.

misrepresented them (Pereboom 2011: 22).

The second case, raised by Pereboom,

involves my daughter, on the occasion of her requiring a Novocain shot at the dentist's. Rather than simply showing her the needle in advance and then giving her the injection, the dentist hid the needle from her and told her that he would be dropping bits of cold water into her mouth. She didn't flinch. When I asked her afterward whether the experience was unpleasant, she said that she didn't like the drops of water much, but they didn't hurt. In this case, it may be that the dentist's suggestion, together with his hiding the needle, kept her from introspectively representing the qualitative features of the pain state she was actually in as qualitative features of pain; instead, she misrepresented those features as qualitative features of a sensation of cold (Pereboom 2011: 23).

There are indeed many possible rejoinders from the friends of introspection. One is that the introspective judgments are most plausibly tainted by doxastic influences intentionally exerted upon the subjects to mislead them, and due to this impurity, these errors do not tell much against the reliability of introspection, as I suggested above, along the line that Hill (1991) himself and many other have suggested.

Here I explore another rejoinder that appeals to PI-R construed under R2. In both cases the conditions of introspection are suboptimal. Moreover, the subjects most likely have not taken, as much as she can, all measures known to be necessary for a reliable exercise of introspection,

in accordance with the guidelines required by PI-R. For the split of second when the student inductee to his fraternity mistakenly believes that he is in pain, his belief is most likely not based on careful introspection. As he pays attention and introspects conscientiously, he comes to the correct belief that it is merely a cold sensation. Similarly, if Pereboom's daughter is indeed in pain, then her belief that she feels none likely results from her inattention induced by the misleading information that her dentist gives her. To my mind it seems plausible that if she had introspected conscientiously, she would have felt pain. However, if it turns out that she would not have felt anything but a cold sensation even upon conscientious introspection, then there is strong reason to say that this is indeed what she actually feels: she never feels any pain and so does not introspect one either. As Schwitzgebel notes in his (2010/2019) survey article, in the contemporary medical literature on pain, no behavioral or physiological markers of pain clearly override self-reports when they come apart (the methodological issue is discussed in Price and Aydede [2005]). Similarly, to my mind there is no clear reason to override the self-report of Pereboom's daughter that she feels no pain – especially if her self-report is based on conscientious introspection – in favor of the belief that her Novocain shot must be painful.

It is true that Pereboom's hypothesis is very modest, as all he intends to establish is an open possibility of introspective inaccuracy. However, to my mind it is still unclear, in the context of conscientious introspection, that this hypothesis represents a serious open possibility: it does not seem any more likely than the radical skeptical hypotheses such as Descartes' (1641/2017) hypothesis that now I am really dreaming and merely seem to be writing this article.

V. Conclusion

By way of conclusion, let us switch gears and look at a moderate take, the most representative of which is Kriegel's (2015) position. Kriegel argues that introspection has both above-chance reliability (ACR) and non-negligible potency (NNP).

ACR: If subject S introspects having phenomenology P, then S is more likely to have P than if S does not so introspect.

NNP: If subject S has phenomenology P, then S is more likely to introspect having P than if S does not have P. (Kriegel 2015: 23).

To me these claims seem true. However, I find them overly concessive. For the present purpose let us focus on the issue of reliability. Why should we retreat to a mere above-chance reliability? Compare the case of perception. When presented with various optical illusions, hallucinations, and other cases problematic for visual perception, it seems to be a hasty overreaction to retreat to the modest claim that our visual perception has a mere above-chance reliability. We need not concede so much in the case of visual perception. Nor need we in the case of introspection.

The non-concessive attitude is especially appropriate given that above I have argued that many concerns with the purported defects of introspection either do not really bear on introspection or have more to do with its impotency than its reliability. The only serious evidence against the reliability of introspection reviewed above does not undermine the reliability of conscientious introspection in ideal conditions. Based on Pelczar's (2014) work, I have defended a principle

of sophisticated introspection that parries off Schwitzgebel's (2011) challenges. Overall the case against introspection is far weaker than the case against visual perception.

WORKS CITED

- Augustine. (420/2003) Bettenson, H. (tr.) *City of God*, London: Penguin Books.
- Ayer, A. (1936/1946) *Language, Truth, and Logic*, London: Gollancz.
- Bayne, T. (2015) Introspective insecurity, in T. Metzinger and J. M. Windt (eds.), *Open MIND: Philosophy and the Mind Sciences in the 21st Century Vol. 1.*, Cambridge: MIT Press.
- Bayne, T. and Spener, M. (2010) Introspective humility, *Philosophical Issues*, **20**, pp. 1-22.
- Berkeley, G. (1710/1734/1982) *Treatise Concerning the Principles of Human Knowledge*, Indianapolis: Hackett.
- Blackmore, S. (2002) There is no stream of consciousness, *Journal of Consciousness Studies*, **9**, No. 5-6, pp. 17-28.
- Byrne, A. (2018) *Transparency and Self-knowledge*, Oxford: Oxford University Press.
- Burge, T. (1996) Our entitlement to self-knowledge, *Proceedings of the Aristotelian Society*, **96**, pp. 91-116.
- Chalmers, D. (2003) The content and epistemology of phenomenal belief, in *Consciousness: New philosophical perspectives*, Smith, Q. and Jokic, A. (eds.), Oxford: Oxford University Press.
- Dennett, D. (1991) *Consciousness Explained*. Boston: Little, Brown & Co.
- Dennett, D. (2002) How could I be wrong? How wrong could I be? *Journal of Consciousness Studies*, **9**, No. 5-6, pp. 13-16.
- Evans, G. (1982) *The Variety of Reference*, Oxford: Clarendon Press.
- Descartes, R. (1641/2017) Cottingham, J. (tr.) *Meditations on First Philosophy: With Selections from the Objections and Replies*, Cambridge: Cambridge University Press.
- Hill, C. (1991) *Sensations: A Defense of Type Materialism*, Cambridge: Cambridge University Press.
- Hohwy, J. (2011) Phenomenal variability and introspective reliability, *Mind and Language*, **26** (3), pp. 261-286.
- Horgan, T. and Kriegel, U. (2007) Phenomenal epistemology: What is consciousness that we

may know it so well? *Philosophical Issues*, **17** (1), pp. 123-144.

Husserl, E. (1913/1982), Klein T. and Pohl, W. (tr.), *Ideas*, Dordrecht: Kluwer.

Kriegel, U. (2015) *The Varieties of Consciousness*, Oxford: Oxford University Press.

Locke, J. (1690/1975) Nidditch, P. (ed.) *An Essay Concerning Human Understanding*, Oxford: Clarendon Press.

Moran, R. (2001) *Authority and Estrangement*, Princeton: Princeton University Press.

Pelczar, M. (2014) *Sensorama: A Phenomenalist Analysis of Spacetime and Its Contents*, Oxford: Oxford University Press.

Pereboom, D. (2011) *Consciousness and the Prospects of Physicalism*, Oxford: Oxford University Press.

Price, D. and Aydede, M. (2005) The experimental use of introspection in the scientific study of pain and its integration with third-person methodologies: the experiential phenomenological approach, in *Pain: New Essays on its Nature and the Methodology of its Study*, Aydede, M. (ed.), Cambridge: MIT Press.

Schwitzgebel, E. (2011) *Perplexities of Consciousness*, Cambridge: MIT Press.

Schwitzgebel, E. (2012) Introspection, what? in *Introspection and Consciousness*, Smithies, D. and Stoljar, D. (eds.), Oxford: Oxford University Press.

Schwitzgebel, E. (2010/2019) Introspection, in *Stanford Encyclopedia of Philosophy*, Zalta, E. (ed.), <<https://plato.stanford.edu/entries/introspection>>.

Siewert, C. (2007) Who's afraid of phenomenological disputes? *Southern Journal of Philosophy*, **45**, pp. 1-21.

Silins, N. (2020) The evil demon inside, *Philosophy and Phenomenological Research*, **100** (2), pp. 325-343.

Smithies, D. (2013) On the unreliability of introspection, *Philosophical Studies*, **165** (3), pp. 1177-86.

Stroud, B. (1984) *The Significance of Philosophical Skepticism*, Oxford: Clarendon Press.