Adriana Renero[1]
and Richard Brown[2]

A HOROR Theory for Introspective Consciousness

Abstract: *Higher-order theories of consciousness typically account for introspection in terms of one's higher-order thoughts being conscious, which would require a third-order thought — i.e. a thought about a thought about a mental state. In this work, we offer an alternative account of introspection that builds on the recent Higher-Order Representation of a Representation (HOROR) theory of phenomenal consciousness. According to HOROR theory, phenomenal consciousness consists in having the right kind of higher-order representation. We claim that this theory can be extended to introspection by recognizing that there is a distinctive kind of consciousness — i.e. introspective consciousness — which can be accounted for as the theory does for phenomenal consciousness generally. We call this novel view: Higher-Order Representation Intentionally For Introspective Consciousness (HORIFIC). We argue that there are independent reasons for thinking that introspective consciousness can be either 'stimuli-induced' or 'self-triggered' and that one of the*

Correspondence:
Email: ARenero@gc.cuny.edu

1 Philosophy Departments, Saul Kripke Center at The Graduate Center, City University of New York, and the Center for Mind, Brain, and Consciousness, New York University, New York, USA.

2 Humanities Department, LaGuardia Community College, City University of New York and MS Program in Cognitive Neuroscience, The Graduate Center, City University of New York, New York, USA.

Journal of Consciousness Studies, **29**, No. 11–12, 2022, pp. 155–73
DOI: 10.53765/20512201.29.11.155

benefits of the view we develop is that it can embrace a pluralist approach. Our view also accounts for what specific mental state is represented by a particular higher-order representation, and for the way in which we are aware of changes, transitions, and boundaries between mental states in specific cases of introspective consciousness.

Keywords: higher-order representation; introspective consciousness; stimuli-induced; self-triggered; Higher-Order Representation Intentionally for Introspective Consciousness.

1. Introduction and Preliminary Considerations

Let us suppose that you are feeling a severe pain in your toe right now. In fact, I am stepping on it while we are standing together. Besides no doubt being in a state of pain, you also feel confused and upset since I am roaring with laughter as I stomp on you. The pain sensation in your toe and the feelings of confusion and upset occupy centre stage in you stream of consciousness seemingly capturing your awareness of such mental states — as occurring to you — without any effort on your part. Take another case. Let us suppose that you have taken your glasses off and right away the text on the computer screen becomes blurry. You are not often tempted to think that the screen — i.e. the physical object — is as such blurry. It is the visual experience of the screen that is blurry, not the screen itself. In such a case you might spontaneously report: 'I cannot see the screen', and be aware of your visual experience.

We take these scenarios to be typical cases of introspective consciousness. In the simplest terms, introspective consciousness is a way to be aware of your own current and recently past mental states, and to self-attribute those mental states to yourself. There seem to be cases where you engage in introspection spontaneously as in the probing of your pain sensation because of my stomping on your foot, *and* cases where you do it purposefully, as when your own interest or volition initiates introspective consciousness of an occurring mental state. In the above example, if I am stomping on your foot to teach you how to introspect painful experiences, then at first you may spontaneously introspectively access the pain sensation. But afterwards you may purposefully direct your powers of introspection just to the pain sensation that you are feeling right now, instead of another state such as a desire to take revenge.

Now, if I ask further how you know that you are having a first-order state — i.e. feeling or experiencing such a stabbing pain or seeing blurry — it is likely that you will express that you have a subjective conviction that — depending on the mental state you are in — you are feeling a stabbing pain in your toe or seeing the screen as blurry. Notice that in asking the question 'how do you know that you are feeling that pain sensation or having such a visual experience?', I am asking for your grounds for judging that you are in that specific mental state. In other words, I am asking what justifies your holding such judgments or self-attributions.[3] Further enquiring like this may require reflection or deeper degrees of introspection, maybe a third-order state, which do not occur regularly or frequently — issues that we leave for another discussion.

What the previous cases suggest is that introspection can be of two different modes: either 'stimuli-induced' or 'self-triggered' (Renero, 2019). A severe pain caused by my stomping on your foot will lead you to introspectively access your pain sensation in a seemingly involuntary way. Here, an episode of introspective consciousness is induced by the stimuli or the pain itself — we shall call this *stimuli-induced introspective consciousness*. In the other case, you voluntarily engage in probing your own mental state occurring. Here, an episode of introspective consciousness is initiated in a voluntary way by selecting a mental state to focus on: either your pain sensation or your emotional response to it, instead of your visual state of your swollen toe, for example — we shall call this *self-triggered introspective consciousness*.

Our goal in this work is to provide the grounds for a novel theory of introspective consciousness that can allow for both of these modes of introspective access *and* to account for the way in which we are aware of changes, transitions, and contrasts between mental states in specific cases of introspective consciousness. To do this, we will build on a version of the higher-order theory of consciousness called the Higher-Order Representation of a Representation (HOROR) theory of phenomenal consciousness (Brown, 2015) aligned with a pluralist view on introspection (Renero, 2019; 2017). We call this novel view: *Higher-Order Representation Intentionally For Introspective Consciousness* (HORIFIC).

[3] Whether such judgments are necessarily describable or verbally reported requires a different discussion that we cannot address here.

Although HOROR has been inspired by the well-known higher-order thought theory of consciousness — i.e. the HOT theory (Rosenthal, 1986; 1993; 1997; 2005; see also Gennaro, 1993; 1996; 2004) — and HOROR could be considered a version of the HOT theory, we will not engage here in a detailed comparison or contrast between HOROR and HOT theories (see Berger and Brown, 2021, for a comparison) or on the benefits of higher-order theories over first-order theories (see Lau and Rosenthal, 2011; Brown, Lau and LeDoux, 2019; Brown, LeDoux and Rosenthal, 2021). We take the HOROR theory as an empirical conjecture about the nature of phenomenal consciousness (Brown, 2014; Brown, Lau and LeDoux, 2019). Although the jury is currently out, we think that the higher-order approach to consciousness is a viable contender and should be explored and developed in enough detail that it can face the tribunal of experience.

The rest of the paper is divided into four main sections: §2 provides some reasons for looking beyond the traditional higher-order account of introspection; §3 introduces the basics of HOROR theory which serves as the basis to develop our view; §4 introduces our novel view, HORIFIC, and its main characteristics; §5 highlights some of the merits of HORIFIC by demonstrating this view at work and closes by offering a summary.

2. Moving Beyond Traditional Higher-Order Views of Introspection

In this section we offer some reasons to look beyond the traditional higher-order view to account for introspection. One very general reason for this enquiry is that introspection has been relatively under-explored in the accounts offered by higher-order theories — even though introspection is considered a 'special case of consciousness' and 'a more complex phenomenon' (Rosenthal, 2005, pp. 27–9). Several authors (e.g. Shargel, 2016; Berger, 2017; Carruthers and Gennaro, 2020) seem to take for granted that any higher-order theory of consciousness will adopt the model of Rosenthal (2005) and Gennaro (2012) where introspection consists in having a conscious higher-order thought. The HORIFIC view contributes by exploring and accounting for this special case of consciousness from a higher-order representation perspective.

Another reason, related to the first, is that philosophical work on introspection is often presented exclusively as some kind of 'self-

monitoring' or 'self-scanning' process targeting a current state (Armstrong, 1968/1993; Rosenthal, 2005; Gennaro, 2012), which sets aside some kind of knowledge by 'acquaintance' or direct awareness to one's mental states (Chalmers, 1996; 2004; Gertler, 2011; 2012). Although introspection can be considered a self-monitoring process, we find that there is something to the phenomenology of acquaintance. There is a sense in which it seems to me that I can have a kind of direct access to my experience or to the phenomenal character of my conscious state, and thus become aware of its nature or, at least, some of its properties. We think that this phenomenology of acquaintance can be accounted for by moving from the traditional higher-order thought view to the HOROR view. The HORIFIC view we develop can also capture the phenomenology of introspection as the content of introspective higher-order representations of first-order representations. It is precisely this phenomenology of introspection that the traditional higher-order account of introspection cannot capture. According to the traditional account, having a conscious higher-order thought — i.e. having a third-order thought representing oneself as thinking a thought about one's first-order mental states (*cf.* Rosenthal, 2005, p. 48; Gennaro, 2012, pp. 56–8) — will make it appear from one's point of view that one is having a conscious thought. That is to say, one will experience consciously thinking that one is seeing red. This is just the result of a kind of self-monitoring or self-scanning process. But consciously thinking a thought such as 'I am seeing red' is not the same as experiencing being acquainted with a phenomenal property.

In fact, according to HOT theory, introspection is having conscious thoughts that one is in particular mental states. But those thoughts do not reveal the nature of those states, nor can we infer them from the close relation between introspection and its content (see Rosenthal, 2005, pp. 43–4). We think that the subject can form representations — instead of thoughts — and although we agree that inference is a different mechanism, we think that introspection can reveal something about the nature of one's target mental states. The HORIFIC view contributes by accounting for the nature of the first-order representations or certain properties of the target mental state.

An additional reason to look beyond the traditional higher-order view to account for introspection is that the classic model of introspection on HOT theory takes introspection as deliberately focusing attention to one's mental states or conscious experiences (*ibid.*, see pp. 48, 108–23). On this view, when one is non-introspectively conscious

one is consciously aware of the world, but when one introspects one comes to shift one's attention to one's conscious experience (*ibid*., see pp. 103–4). It may be the case that we can have this kind of deliberate and focused access, and it may fit well with our proposal of a self-triggered introspective consciousness. However, it would not capture or fit well with stimuli-induced introspective consciousness where one's awareness is drawn to the experience, not to a thought about it — this is to say that it seems to one that one is aware of the experience rather than having a conscious thought (see above our second reason).

Furthermore, the HORIFIC view contributes by proposing an account which includes both the stimuli-induced and self-triggered modes of introspective consciousness — a distinction that has been neglected in the philosophical literature on introspection. This account builds on HOROR theory and aligns with a pluralist theory of introspection, it also extends them by advancing proposals specific to this special case of consciousness. The HORIFIC view extends well beyond what is present in work on HOT and HOROR theories of consciousness, and it could be adapted to fit with other theories of consciousness as well.

3. The HOROR Theory of Phenomenal Consciousness

Since our goal is to build upon the HOROR theory of phenomenal consciousness, we will start with a very brief account of the basics of the theory. HOROR theory is a higher-order representational theory of phenomenal consciousness. *Phenomenal consciousness*, in the most general sense, is just the property a creature has of there being something that 'it is like' for them (Nagel, 1974). Conscious experiences are distinguished from each other by their specific phenomenal character. In this sense, the *phenomenal character* of an experience is just the specific way that the conscious experience is like for the creature in question — e.g. consciously seeing blue, hearing a trumpet, thinking that 'cinnamon' is hard to spell, etc. This notion of phenomenal consciousness is neutral and accepted by theorists with different approaches to consciousness (e.g. Chalmers, 2018; Raccah, Block and Fox, 2021). 'Phenomenal character' captures the specific property in which conscious experiences differ from each other. A visual experience of seeing blue versus seeing green, and an auditory experience of hearing a trumpet versus hearing a double bass, will differ in their phenomenal character. For one it will be like seeing the

colour quality of blue or green, and the other will be like hearing the sound quality of a bright-tone trumpet or the rustling-like sound of a double bass.[4]

Providing an account of phenomenal consciousness is the primary goal of any theory of consciousness. One approach to understanding the nature of phenomenal consciousness is via *representationalism*. In its simplest form, representationalism is the view that holds that phenomenal consciousness supervenes on, or is identical with, some kind of impure representation; where that means a representational content represented in some particular way. Representationalism comes in two kinds: first-order and higher-order. On the one hand, first-order theories (e.g. Tye, 1994; Dretske, 1995) hold that phenomenal consciousness consists in representations of properties in the environment — though see Gottlieb (2019) for an argument that these theories collapse into higher-order theories. On the other hand, higher-order theories hold that phenomenal consciousness crucially involves representations of one's own mental life. Since we are typically aware of things either by perceiving them or by thinking that they are currently present, higher-order theories have been divided into higher-order *perception* (e.g. Armstrong, 1968/1993; Lau, 2019) and higher-order *thought* (Rosenthal, 2005) theories.

The HOROR theory starts with the folk platitude that consciousness involves an awareness of our mental life and then identifies the appropriate kind of awareness with an appropriate higher-order *representation*. Here 'appropriate' means that the higher-order representation subjectively appears to have arisen spontaneously, independently of any inference and that it represents oneself as currently being in the *target* mental state. It also postulates that the contents of the higher-order representations will account for the phenomenal character of one's experience.[5] When one consciously sees, for example, the blue sky, one would then be aware of oneself as seeing blue. One would

4 For a view of consciousness which includes both an account of the phenomenal character of mental states and how those states become conscious — i.e. a view that builds on the HOT theory of consciousness plus the quality-space theory of mental qualities (and how these theories can work in tandem), see Renero (2014).

5 Here we use the term *target* mental state to designate the mental state which is represented by the relevant higher-order representation. We use 'content' of a representation to refer to the way in which the representation in question represents its target. If I am in a state which is a representation of something non-mental, then the target is the physical object and the content amounts to the satisfaction conditions placed on the targeted object.

attribute to oneself an occurrent mental state, say, a representation of seeing blue. In a nutshell, HOROR theory is a representational theory in that it holds that phenomenal consciousness can be understood in terms of representational states. Specifically, the HOROR theory says that when one is phenomenally conscious, one has the appropriate kind of higher-order representation.

We can put the HOROR theory of phenomenal consciousness more specifically:

> **HOROR:** For a subject *S* to be in a phenomenally conscious state *C* with phenomenal character *P* is just for *S* to token an appropriate higher-order representation with the content that *S* is in *C*, which has character *P*.

As we can see, the basic idea is that, for phenomenal character, if one represents oneself as having it, then one does in fact have it. Let us suppose that *S* is looking at a ripe tomato right now, so *S* has a phenomenally conscious state. The phenomenal character of *S*'s visual state of the tomato is its redness and roundness. Notice that while a first-order representational theory will account for this in terms of a first-order representation of the redness and roundness of the tomato, a higher-order representation view will account for this in terms of a higher-order representation of oneself as being in those kinds of first-order representational states.

In this case, the suitable higher-order representation will have as its content something like the following: *I am seeing a red round object*. That is, of course, a rough approximation of the content. But the basic idea is that the content of the higher-order representation deploys concepts which describe one as oneself currently *being in* the representations which are targeted by the appropriate higher-order representations of a representation (HORORs), and so it will seem from one's point of view that one *is* in those mental states. The targeted first-order representations are characterized by the targeting HOROR as presenting objects in the environment which have properties like colour, sound, etc. The HOROR theory is built with this commonsense picture in mind and it is this pre-theoretic notion that the colours, sounds, etc. seem to be out there on the objects or events that the theory aims to account for. One does this, according to the HOROR theory, by representing oneself as being in states that put you in a special relation to perceptible properties. That is all there is to

being phenomenally consciousness according to the HOROR theory.[6] When one has a HOROR like this, one will be phenomenally conscious and will experience the blue sky, the ripe tomato, the bright-tone trumpet, or whatever it is that such a HOROR says you are currently representing. Notice that this is not to say that the world *is* as we experience it! But the important point is that HOROR theory vindicates the common-sense idea that in consciously experiencing the blueness of the sky, the redness of the tomato, or the brightness-tone of the trumpet, it does not necessarily seem to me as though I am aware of some mental quality. I seem to be aware of the tomato itself and its redness. According to the HOROR theory, this is because the higher-order representations describe oneself as being in states which present one with objects in the environment that have colours, sounds, etc. and so this is how one experiences the world.

4. Higher-Order Representation Intentionally For Introspective Consciousness (HORIFIC)

What is the HORIFIC theory? When one has a conscious experience of seeing blue and one then introspects such a mental state, one still is visually experiencing blue but one is also aware of the blueness as a property of one's visual experience. 'What it is like' for one is *like* consciously probing and focusing to one's own mental life. When one introspects a conscious experience, one is in a special relationship with one's conscious experience.[7] A brief comparison is worthwhile: While HOROR is a theory of phenomenal consciousness which accounts for the character of the experience as a property of — or related to — an object of the external world, the HORIFIC theory extends to introspective consciousness and builds upon the HOROR theory. The HORIFIC theory accounts for the distinctive way in

[6] One reason for adopting HOROR theory comes from Block's (2011) attack on Rosenthal's HOT theory. If there is phenomenal consciousness in the absence of the *first-order state*, then we seem to have a conscious state with no neural correlate. Rosenthal seems to agree that the correlate of the conscious experience is the correlate of the *higher-order representation*. So, we see him as ultimately agreeing with us about HOROR. Further discussion and analysis must await another occasion.

[7] Advocates of the transparency of experience may deny these claims about introspection; defending them is beyond the scope of this article. Here, we note that the kind of examples presented at the beginning of the paper involving taking off one's glasses suggest that one can become aware of one's experience. This suggests that the strong claim that we are only ever aware of properties of external objects and never aware of any of our mental properties is false.

which the relevant experience appears *to oneself* as a property of — or related to — one's own experience or occurrent mental state.

We can put the HORIFIC theory of introspective consciousness more specifically:

> **HORIFIC:** For a subject *S* to be introspectively conscious of a mental state *m* as having character *Q* is for *S* to token an appropriate higher-order representation with content that *S* is experiencing *m* with character *Q*.

The HORIFIC theory, then, accounts for the distinctive way in which a first-order representation is conscious to the subject. A caveat here is important. For both HOROR and HORIFIC theories, the distinctive 'what-it-is-like-ness' of an experience is exhausted by the representational content of the appropriate higher-order representation. However, different versions of the HORIFIC theory may postulate different contents depending on how they characterize the term introspective consciousness. As for now, we have not developed other versions of the theory.

Suppose one is looking at a basket of ripe strawberries which are deeply red and looking delicious. As we have seen, according to the HOROR theory one's experience of the redness and fragrance, for example, consists entirely in tokening an appropriate higher-order representation attributing to oneself being in states which represent the strawberries. But what happens when one becomes introspectively conscious? According to the HORIFIC view, one comes to form an introspective HOROR. That is, one comes to form an introspective representation of the same first-order representations one was already targeting. The targeted first-order representations are themselves mental states which represent the strawberries — and which have their own functional roles. These states may involve mental qualities, concepts, or whatever it is that one thinks allows the creature in question to be aware of the properties in the environment that matter to it. Let us call this collection of first-order representations *m* and the original HOROR *C* (for consciousness). Then, *C* targets *m* and represents oneself as being in *m.* When one then introspects, *C* can come to be replaced by *I* — i.e. an introspective HOROR — where *I* targets the very same first-order representations *m* and represents oneself as experiencing being in them.

The introspective HOROR is *not* an additional higher-order representation in the sense of a third-order representation of the original HOROR *C* from above. Rather, it is just a *different* second-order

representation that targets the very same first-order perceptions but with a new introspective content. Introspective consciousness is a kind of phenomenal consciousness — a probed and focused version of ordinary non-introspective consciousness — and the HORIFIC view explains it in the same way as it does phenomenal consciousness generally.

On the HORIFIC view, the higher-order representation represents oneself as experiencing the first-order representations. The content of this introspective higher-order representation, roughly, would be that *I am experiencing — perceptible — red in a distinctly visual way*. Since this introspective representation is a HOROR, it follows (*ex hypothesi*) that the subject has an experience which is exhaustively characterized by the representational content of the HOROR. In this case one would be *experiencing seeing red*. This is what it is to introspect a state about the redness of the strawberries on the HORIFIC view.

5. HORIFIC: A Pluralist View about Modes of Introspective Consciousness

If the HORIFIC theory is true, there is room to account for two different modes of introspective consciousness *and* for the nature of the first-order representation at issue — or at least certain properties of the target mental state.

As for the former claim, if introspective consciousness is considered of two modes, this will open up the possibility of postulating a pluralist approach, either:

> **Stimuli-induced:** A receptive introspective consciousness whereby a specific mental state or first-order state spontaneously or automatically induces a higher-order representation (see Renero, 2019, p. 837).
> **Self-triggered:** A selective introspective consciousness whereby the subject's own interest and volition initiate a higher-order representation of an occurrent first-order representation (*ibid.*, see p. 840).[8]

[8] It is important to clarify that we do not mean here having any voluntary control over the mental state or a capacity to inhibit the higher-order representation at issue. But we do refer to being able to be introspectively conscious of what is there to be represented. In this sense, there is an active role being undertaken by the subject in this particular mode of introspective consciousness.

On the one hand, in considering *stimuli-induced* introspective consciousness, it is important to see that the stimulus is not necessarily caused or induced by an external source of the physical world but is a (first-order) mental-state-induced representation. Stimuli-induced introspective consciousness shows that the character of the mental stimulus induces different representations and exhibits distinct outputs, which can be either simple representations or complex representations that emerge, such as when you hit me: I respond with pain, I shout or express anger, I hit you back, or run away.

Following our initial case of (a target mental state such as) a pain sensation, consider an example of a *simple output*: a first-order state such as a severe pain sensation in your toe induces an introspective HOROR of that state and you describe such a state as a stabbing: 'I'm experiencing having a stabbing sensation' — after ruling out a throbbing sensation, for example. Let us call it a higher-order representation induced by a painful sensation.

Now, consider an example of a *complex output*: a first-order state such as a severe pain sensation in your toe induces a higher-order representation of that state, which brings about a new target state: a visual state of your swollen toe. And this state may also induce a simultaneous mental state: a deep concern about paying for a medical consultation to get an examination of your toe, or a real desire that I — the tormentor who has stomped on you — disappear instantly. Accordingly, you describe such a state as a swollen toe: 'I'm experiencing having a swollen toe, now I need to see a doctor, I do not have health insurance and I'm bankrupt.'

On the other hand, in considering *self-triggered* introspective consciousness, it is important to see a different scenario. First, take the following context: as always, you are rushing. But you prefer to stop for a minute to enquire as to what your current first-order state is. You choose to start a self-probing of your own mental life and form a higher-order representation of your current mental state — i.e. one first-order state that you select from a cluster of occurring first-order states as they appear in your mind. Now, consider the previous pain sensation example as happening to you right now: you form a higher-order representation of your current pain sensation — and set aside for the moment other first-order states that you can generate — e.g. any emotional response to it, or your desire to take revenge. Self-triggered introspective consciousness is accompanied by an intention to undertake a search, or by simple inquisitiveness as one self-probes one's first-order mental representation for the sake of investigation.

As mentioned, adopting the possibility to account for distinct modes of introspective consciousness — stimuli-induced or self-triggered — promotes a pluralist approach. The HORIFIC theory is pluralist since introspective consciousness would not be restricted to a single relation, nor would it be reducible to a unique form of access or awareness of one's mental states.[9] Instead, we propose that higher-order representations can represent first-order mental states in distinct modes. Those modes may vary depending on time and specific situation.[10] While stimuli-induced introspection may exhibit distinct types of outputs, self-triggered introspection may reveal the selected mental state in various ways and furnishes the experience. A pluralist approach such as this allows the possibility of broadening the notion of introspective consciousness while other theories of introspection either neglect it or maintain constraints.[11] We will not engage here in an examination of these theories, or in a comparison or contrast between HORIFIC and other approaches.[12] The relevant point here is that this novel consideration had also been left out by leading views of consciousness.[13]

9 One can also be a pluralist concerning the variety of the target mental states that are represented via HORORs. That is, we can have higher-order representations that target *all* types of target mental states — i.e. *cognitive states* or propositional attitudes such as beliefs and *non-cognitive states* such as sensations and perceptions or those experiences that are linked to our sensory modalities (seeing, hearing, smelling, tasting, and touching); and our affective states or those experiences such as emotions, feelings, and bodily sensations like pains, itches, tingles, and cramps — or we can have higher-order representations that target only *some* of them. Although we think that HORIFIC can represent *all* kinds of states as its *targets*, offering an account of this possibility goes beyond the scope of this paper.

10 Notice that an introspective episode can begin by self-triggered introspective consciousness and end up with a verbal report, or it can begin by stimuli-induced introspective consciousness and end up solely in a silent judgment.

11 For example, single theories such as the *inner-sense* view account for a mere causal or mediated relation between the target mental state and the introspection of such a target. Whereas theories such as the *acquaintance* view account for a direct or immediate relation between the target mental state and introspection of such a target. These theories claim that only one type of awareness, either the causal one or the direct one, is defined as introspection. Presenting a non-exclusive alternative view which can accommodate both relations may be more attractive than other restrictive positions. For a discussion between these views and a possible conciliation, see Renero (2019).

12 Exploring whether the HORIFIC theory is compatible with the mentioned views requires a different analysis that goes beyond the present purview.

13 For example, HOT theory is usually considered a *single* theory — as opposed to a pluralist theory — in which the higher-order thought is caused by a stimuli of the external world. HOT theory is also identified among the inner-sense views similar to

Although the fundamental differences between stimuli-induced and a self-triggered introspective consciousness are clearly seen in the definitions and characterizations above, close attention to the previous scenarios reveals that introspective consciousness — in either form — fulfils three criteria, at least: 'it is directed at one's mind (first-person); it is about psychological states; namely, mental entities, as opposed to non-mental entities (mental); and it is about one's current, ongoing, and recent past mental states (occurring)' (Renero, 2019, p. 824). Now, introspective individuation of one's occurring states or experiences is given as a function of these criteria plus other conditions of access such as consideration of context, possession of right cognitive capacities, or acquisition of the relevant conceptual resources to describe or report one's own mental states (*ibid*., see p. 833). These criteria are presupposed in the examples that we have offered here.[14]

Let us turn to the latter claim — i.e. that the HORIFIC theory can also tell us something about the nature of the first-order target mental state at issue. That is, the HORIFIC view can account for which specific mental state is the target of a particular higher-order representation — e.g. whether my first-order representation is a pain sensation versus a tickle, or whether it is a desire rather than a whim given its particular phenomenal character — *and* for the content or the way in which one represents oneself as having the experience and how one can eventually describe the relevant first-order state in terms of *changes*, *transitions*, and *boundaries* between that state and other occurring states in accordance with time and specific situation.

As for the nature of the first-order representation or certain properties of the target mental state, we contend that:

P1. Subject *S* can introspectively individuate mental states according to their properties as changing or *varying*, transitioning or *passing through*, and dividing or *delimiting*, while they are appearing to *S*'s mind and *S* forms a corresponding higher-order representation.

Armstrong's view (1968/1993). Discussion about this particular issue is left for future work.

14 Although we acknowledge that offering analysis on the metaphysics of experiences is relevant and necessary for further justification, we will have to leave this analysis for future work.

P2. If P1 holds, the HORIFIC theory accounts for introspective consciousness as a dynamic and flexible phenomenon and HORIFIC can be a promising theory.

———

C. The HORIFIC theory accounts for introspective consciousness as a dynamic and flexible phenomenon and HORIFIC can be a promising theory.

First, concerning *changes*, consider (a) the change or shift of direction from a recently past mental state to a current mental state within a specific introspectively conscious event (ms_1 at t^1, ms_2 at t^2), or (b) the *change* — which can be sudden or gradual — in phenomenal character of an experience within a specific introspectively conscious event (ep_1 at t^1, ep_2 at t^2). For example, (a) while you are forming a higher-order representation of your visual perception of a bottle of mezcal, from the phenomenal character of this visual experience you can be introspectively conscious of a sudden change in your target mental state: your visual perception has just shifted into the desire to have a glass — or two! — of mezcal. Notice here that the type of target mental state at t^1 changes to another target mental state at t^2. Alternatively, (b) while you are probing your gustatory sensation of the mezcal and enjoying its strong smoky flavour, you find out by introspective consciousness that the character of your experience suddenly changes when you see a worm inside the bottle. Notice here that the phenomenal character of the target mental state occurring at t^1 changes given a new content of the representation at t^2.

Second, concerning *transitions*, consider the transition from an occurring mental state to a fading or dissipating mental state within a specific introspectively conscious event (from ms_1 to ms_0 at t^1), or the pathway from having a powerful experience to including it within an introspectively conscious event (from e_1 to e_0 at t^1). For example, while you are probing your auditory state of hearing yourself laugh and enjoying the bodily sensation that it brings about, you detect by introspective consciousness the transition in the state from a chuckle to a smile, to a feeling of embarrassment; and all this precisely at the time you notice you are laughing with a horrible screeching sound.

Third, concerning *boundaries* which divide or delimit one mental state from another according to its duration or scope,[15] consider the

[15] Even if the introspective event is relatively short, its duration is susceptible to further extension if introspective consciousness continues as it might (or might not).

boundary or the limit between (a) two joint mental states occurring at the same time within a specific introspectively conscious event (ms_1 and ms_1* at t^1), or (b) two similar or phenomenally related mental states having different content within a specific introspective event (ms_1 about *s* and ms_1 about *d*). For example, (a) while you are probing your auditory sensation, which you savour, of simultaneously (audible) crying and laughing, you detect by introspective consciousness the boundaries between this particular auditory experience — usually characterized by a series of spontaneous sounds — and the delightful experience that accompanies it. Although these states occur at the same time, they have different duration or scope. Either that, or (b) while you are self-probing your mental state of sadness because of a particular loss, you discover by introspective consciousness the boundary between this sadness expressed by crying is, on the one hand — or in the beginning — represented as sorrowful, and, on the other hand — or when it is coming to a close — represented as desperate.

Boundaries are particularly helpful in recognizing how instances of target mental states of shortened duration occur, how repetitions of a certain mental state come along, and also in delimiting when exactly the target mental state starts and ends, when an earlier mental state is similar to a current state, and so forth. Thus, we say that one is able to form corresponding higher-order representations of one as having those particular experiences. The formation or generation of higher-order representations according to changes, transitions, and boundaries is given as a function of their corresponding first-order states — i.e. in how many distinctions in her conscious experiences the introspective subject can attain. So, we say that the introspective subject will be having higher-order representations that assert that she is in a particular mental state with certain phenomenal character and content. Differences in phenomenal character are reflected in the intentional content of her representations, which are themselves more refined in descriptions or verbal reports.

It is relevant for our theory to emphasize that the introspective subject is aware of being in certain first-order mental states as they change sequentially, the transition between themselves, or the boundaries they maintain. These introspective experiences, on HORIFIC, amount to the changing contents of the relevant higher-order states, which themselves are driven or partially driven by the causal connections to the first-order states themselves (stimuli-induced) and/or initiated or partially initiated by the own interest and volition of the subject who is introspecting (self-triggered). Those higher-order states

can also be combined such as when introspective consciousness is driven by a stimulus and, then, the output demands self-triggered or further introspective consciousness about the target mental state.[16] These considerations involve the possibility of accounting for higher-order representations in accordance with the mentioned two different modes and depending on the property or the content of the first-order state to be represented.

On the HORIFIC view, introspective consciousness is a flexible and dynamic phenomenon; it can respond to the changes of one's experiences according to their phenomenal character and content, while identifying and distinguishing other accompanying mental states. Introspective consciousness is also responsive to fluctuating mental stimuli by its flexibility to switch from one mental state at t^1 to another mental state at t^2 and, accordingly, to form different higher-order representations. The same applies to transitions and boundaries between the mental states that accompany other higher-order representations.

To close this section, let us illustrate the core of the HORIFIC view to show what this theory accounts for in the following diagram (details and examples have been offered above).

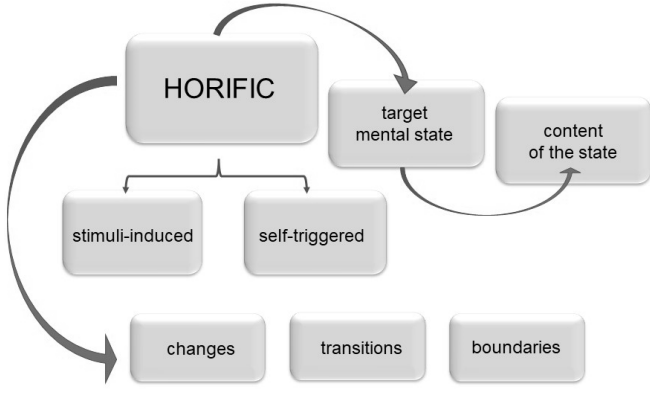

Figure 1. Introspective consciousness can be receptive (*stimuli-induced*) or selective (*self-triggered*). A higher-order representation for introspective consciousness can reveal the mental state which is represented (the *target* state) and the way in which the target state is represented (the *content* of such a state). It can also distinguish *changes*, *transitions*, and *boundaries* between mental states in certain episodes of introspective consciousness.

16 Whether both modes of introspective consciousness can occur together regarding the same mental state is a subject that remains to be worked out in future research.

6. Conclusion

We have offered the basics of the HOROR theory of phenomenal consciousness and have built upon it to extend it to introspective consciousness via the grounds of — what we have called — the HORIFIC theory. The relevant point is that one would be able to form a higher-order representation of a first-order representation — i.e. the target mental state — and thus be introspectively conscious of such a representation. Having highlighted the merits of HORIFIC by demonstrating this view at work, we have further shown that some of the theoretical consequences of holding this novel view entail the possibility of providing a pluralist perspective: the target mental states of higher-order representations are not formed just in one way or represented in the same way. Rather, one can represent first-order states and certain properties of those states based on one's own experience of certain stimuli, or one can represent first-order states and certain properties of those states based on one's own interest and volition to initiate a self-probing process. This is a promising theory for investigating the relevance of higher-order approaches in connection to introspection, while paving the way for further research on introspective consciousness.

References

Armstrong, D.M. (1968/1993) *A Materialist Theory of the Mind*, New York: Humanities Press.

Berger, J. (2017) How things seem to higher-order thought theorists, *Dialogue*, **56** (3), pp. 503–526.

Berger, J. & Brown, R. (2021) Conceptualizing consciousness, *Philosophical Psychology*, **34** (5), pp. 637–659.

Block, N. (2011) The higher-order approach to consciousness is defunct, *Analysis*, **71** (3), pp. 419–431.

Brown, R. (2014) Consciousness doesn't overflow cognition, *Frontiers in Psychology*, **5** (1399). doi: 10.3389/fpsyg.2014.01399

Brown, R. (2015) The HOROR theory of phenomenal consciousness, *Philosophical Studies*, **172 (**7), pp. 1783–1794.

Brown, R., Lau, H. & LeDoux, J.E. (2019) Understanding the higher-order approach to consciousness, *Trends in Cognitive Sciences*, **23** (9), pp. 754–768.

Brown, R., LeDoux, J.E. & Rosenthal, D.M. (2021) The extra ingredient, *Biology and Philosophy*, **36** (2), pp. 1–4.

Carruthers, P. & Gennaro, R. (2020) Higher-order theories of consciousness, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy*, Fall 2020 Edition, [Online], https://plato.stanford.edu/archives/fall2020/entries/consciousness-higher/ [22 April 2022].

Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.

Chalmers, D.J. (2004) How can we construct a science of consciousness?, in Gazzaniga M.S. (ed.) *The Cognitive Neurosciences III*, pp. 1111–1119, Cambridge, MA: MIT Press.
Chalmers, D.J. (2018) The meta-problem of consciousness, *Journal of Consciousness Studies*, **25** (9–10), pp. 6–61.
Dretske, F. (1995) *Naturalizing the Mind*, Cambridge, MA: Bradford Books.
Gennaro, R.J. (1993) Brute experience and the higher-order thought theory of consciousness, *Philosophical Papers*, **22**, pp. 51–69.
Gennaro, R.J. (1996) *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*, Amsterdam: John Benjamins.
Gennaro, R.J. (2004) Higher-order thoughts, animal consciousness, and misrepresentation: A reply to Carruthers and Levine, in Gennaro, R.J. (ed.) *Higher-Order Theories of Consciousness*, Amsterdam: John Benjamins.
Gennaro, R.J. (2012) *The Consciousness Paradox*, Cambridge, MA: MIT Press.
Gertler, B. (2011) *Self-Knowledge*, New York: Routledge.
Gertler, B. (2012) Renewed acquaintance, in Smithies, D. & Stoljar, D. (eds.) *Introspection and Consciousness*, New York: Oxford University Press.
Gottlieb, J. (2019) The collapse argument, *Philosophical Studies*, **176**, pp. 1–20.
Lau, H. (2019) Consciousness, metacognition, and perceptual reality monitoring, *PsyArXiv*. doi: 10.31234/osf.io/ckbyf
Lau, H. & Rosenthal, D. (2011) Empirical support for higher-order theories of conscious awareness, *Trends in Cognitive Sciences*, **15**, pp. 365–373.
Nagel, T. (1974) What is it like to be a bat?, *Philosophical Review*, **83** (4), pp. 435–456.
Raccah, O., Block, N. & Fox, K. (2021) Does the prefrontal cortex play an essential role in consciousness? Insights from intracranial electrical stimulation of the human brain, *The Journal of Neuroscience*, **41** (10), pp. 2076–2087. doi: 10.1523/jneurosci.1141-20.2020
Renero, A. (2014) Consciousness and mental qualities for auditory sensations, *Journal of Consciousness Studies*, **21** (9–10), pp. 179–204. doi: 10.53765/20512201.21.9.179
Renero, A. (2017) The nature of introspection, *CUNY Academic Works*, [Online], https://academicworks.cuny.edu/gc_etds/2144.
Renero, A. (2019) Modes of introspective access: A pluralist approach, *Philosophia*, **47**, pp. 823–844. doi: 10.1007/s11406-018-9989-2
Rosenthal, D.M. (1986) Two concepts of consciousness, *Philosophical Studies*, **49**, pp. 329–359.
Rosenthal, D.M. (1993) Thinking that one thinks, in Davies, M. & Humphreys, G. (eds.) *Consciousness*, Cambridge, MA: Blackwell.
Rosenthal, D.M. (1997) A theory of consciousness, in Block, N., Flanagan, O. & Güzeldere, G. (eds.) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press.
Rosenthal, D.M. (2005) *Consciousness and Mind*, New York: Oxford University Press.
Shargel, D. (2016) The insignificance of empty higher-order thoughts, *Journal of Cognition and Neuroethics*, **4** (1), pp. 113–127.
Tye, M. (1994) Qualia, content, and the inverted spectrum, *Noûs*, **28**, pp. 159–183. doi: 10.2307/2216047

Paper received November 2021; revised June 2022.