Received: 1 October 2019 | Revised: 2 March 2020 | Accepted: 30 March 2020

DOI: 10.1111/mila.12363

SUBMITTED ARTICLE

WILEY

Underwhelming force: Evaluating the neuropsychological evidence for higher-order theories of consciousness

Benjamin Kozuch

Department of Philosophy, University of Alabama, Tuscaloosa, Alabama

Correspondence
Benjamin Kozuch, Department of Philosophy, University of Alabama, 336 ten Hoor Hall, 350 Marrs Spring Road, Tuscaloosa, AL 35401.
Email: bkozuch@ua.edu

Funding information
College Academy of Research, Scholarship, and Creative Activity, University of Alabama, Grant/Award Number: 11010-204109-100; University of Alabama Research Grants Council, Grant/Award Number: RGC-2018-55

Proponents of the higher-order (HO) theory of consciousness (e.g., Lau and Rosenthal) have recently appealed to brain lesion evidence to support their thesis that mental states are conscious when and only when represented by other mental states. This article argues that this evidence fails to support HO theory, doing this by first determining what kinds of conscious deficit should result when HO state-producing areas are damaged, then arguing that these kinds of deficit do not occur in the studies to which HO theorists appeal. The article also develops an apparatus that can be used to evaluate whether other lesion evidence confirms or disconfirms HO theory.

KEYWORDS
brain lesions, consciousness, higher-order theories of consciousness, prefrontal cortex, vision

1 | INTRODUCTION

One currently popular approach to understanding consciousness is the *higher-order* theory of consciousness, which hypothesizes that a mental state is conscious if and only if it is represented by another, certain kind of mental state (e.g., Carruthers, 2000; Kriegel, 2009; Lycan, 1996; Rosenthal, 2002).[1] The theory is at least initially plausible: Consider how it would be odd to say that some mental state is conscious, were the subject entirely unaware of it. This gives us the

[1]Other philosophers or psychologists subscribing to higher-order theory (if not always by name) include Locke (1690), Armstrong (1968), Gennaro (1996), Rolls (2004), Van Gulick (2004), Lau (2008b), Brown (2015), and LeDoux (LeDoux & Brown, 2017).

idea that conscious mental states are those mental states that the subject is aware of being in, the so-called *transitivity principle* (Lycan, 2001; Rosenthal, 1997). If one takes awareness to be a matter of representation (Lycan, 2001), then we arrive at what we might call the "higher-order thesis," which says that a mental state is (phenomenally)[2] conscious if and only if it is represented by another, certain kind of mental state.[3] While there has been debate as to what exactly the right kind of higher-order state is (e.g., whether it is a "perception" or a "thought"),[4] higher-order theorists of all types agree that *some* kind of higher-order representation is necessary and sufficient for a mental state being conscious.

The lively and protracted debate brought about by higher order theory (hereafter, HO theory) has mostly proceeded through a priori argumentation, largely insulated from empirical data. In the last decade, however, the debate has been re-conceptualized as one to be settled through neuroscience, with numerous articles taking this approach recently appearing.[5] Typically, a two-step strategy is used: First, it is hypothesized that certain brain areas are candidates for producing the kind of HO state necessary and sufficient for having a conscious state; second, neuroscientific data are used to argue that activity in these brain areas is or is not somehow essential for conscious experience, where this counts as evidence for or against HO theory, respectively.[6] Using this strategy requires having some idea as to which brain areas are candidates for producing the right kind of HO state. (Henceforth, the phrase "the right kind of" is mostly left implied). A popular candidate for containing such areas is the part of the brain where many metacognitive functions are carried out, the *prefrontal cortex* (PFC) (Metcalfe & Schwartz, 2015).

Now, if HO state-producing areas were located in the PFC, it seems that a prediction of HO theory would be that PFC lesions will sometimes cause deficits in consciousness. However, a number of commentators have used evidence in which PFC lesions appear to not lead to such deficits to argue that activity in the PFC is not essential for consciousness (Pollen, 2008; Gennaro, 2012; Kozuch, 2014; Boly et al., 2017; but see Odegaard, Knight & Lau, 2017). To counter such data, HO theorists (e.g., Lau & Rosenthal, 2011a; Lew & Lau, 2017; Morales & Lau, 2020) have cited studies in which PFC lesions seem to cause conscious visual deficits. Remarkably, these very same studies have sometimes been used to argue *against* HO theory (Kozuch, 2014).

The disagreement here seems to be one concerning what kinds of conscious deficit we should expect when an HO state-producing area is lesioned. Perhaps it is not surprising that this disagreement is unresolved, since we are yet to have an in-depth treatment of this issue. Given this, a primary goal of this article is to fill this gap in the literature, providing a better developed account of what should phenomenologically result when HO state-producing areas

[2]Here, I am concerned with higher-order theory insofar as it is meant to be a theory of phenomenal consciousness, understood as the "what-it's-like" aspect that some mental states have (Block, 1995; Chalmers, 1995; Nagel, 1974).

[3]Not everyone finds this reasoning convincing (see, for example, Block, 2009).

[4]The distinction here would be between higher-order *perception* (Armstrong, 1968; Lycan, 1996) and higher-order *thought* (Carruthers, 2000; Rosenthal, 2002) theories. Other variations include whether the theory takes the higher-order and lower-order states to be numerically distinct (Lycan, 1996; Rosenthal, 2002) or identical (Carruthers, 2000; Gennaro, 1996; Kriegel, 2009; Van Gulick, 2004), and whether a mental state's becoming conscious requires an HO state actually targeting it (e.g., Rosenthal, 2002), or there only be a disposition for this to occur (Carruthers, 2000).

[5]Examples include Kriegel (2007, 2009), Lau (2008b), Lau and Rosenthal (2011), Gennaro (2012, 2016), Kozuch (2014), Sebastián (2013), Lau and Brown (2019), and Block (2019).

[6]Sebastián (2013), for instance, appeals to masking studies as evidence for the dorsolateral PFC (dlPFC) producing HO states, then argues against HO theory by using data showing reduced dlPFC activity during periods of dreamful sleep (i.e., when the subject is having the conscious experiences associated with dreaming).

are damaged. Accomplishing this first goal will help to fulfill the other primary goal of this article, which is to determine whether the PFC studies to which HO theorists appeal actually support HO theory. Accomplishing this first goal will also help to evaluate arguments that have used PFC lesion data as evidence against HO theory (Kozuch, 2014) or against the PFC being essential for consciousness (e.g., Pollen, 2008, Boly et al., 2017), as well as similar arguments to be made in the future.

Here is the article's layout: Section 2 discusses how lesion data could be used to support HO theory, also examining the lesion studies to which HO theorists have appealed to support their theory. Section 3 investigates the issue of what should result when HO state-producing areas are impaired, arguing that this would likely cause a large number of visually conscious states to be lost. Section 4 uses this investigation's results to argue that, in the studies to which HO theorists appeal, the lesions have not caused the loss of many visually conscious states, and that these studies therefore do not confirm HO theory. Section 4 also discusses the issue of whether the studies to which HO theorists appeal (or other prefrontal lesion studies) could be considered to *dis*confirm HO theory.

2 | BACKGROUND: USING LESION STUDIES TO INVESTIGATE HO THEORIES

HO theory entails certain neuroscientific predictions concerning lesions, namely that PFC lesions should result in some kind of conscious deficits. This provides a way to confirm HO theory using lesion data. This section examines how lesion studies have been used to try to support HO theory, also taking a detailed look at some of the studies to which HO theorists have appealed.

2.1 | HO theory and lesion evidence

There is a profitable framework within which to understand how lesion studies can be used to confirm HO theory (Kozuch, 2014), which looks as follows: Consider that, if HO theory is true, and HO states are needed for having conscious states, then there must be some brain area (or areas, or network of areas),[7] such that, because it produces these HO states, its operating properly is necessary for conscious experience. We can refer to this as an "integral area." From the idea of an integral area follows a neuroscientific prediction, which is that, when an integral area is damaged, this could cause deficits in consciousness, since it could diminish one's ability to produce HO states. But the concept of an integral area does not allow us to generate testable predictions, since we do not yet know which brain areas, if any, are integral. But consider that there might be areas in the brain such that, because we have reason to believe that they might produce HO states, they count as *potential* integral areas. This does lead to a testable prediction: If HO theory is true, then lesions to potential integral areas are likely to sometimes cause

[7]The parenthetical statement recognizes the possibility (a) that there is more than one brain area individually sufficient for producing HO states or (b) that producing HO states might require the participation of more than one brain area (i.e., a network of areas). Both possibilities are taken into account below (see especially Section 3.1; for discussion, see Kozuch, 2014).

deficits in consciousness. It seems then that, if there is an instance where potential integral area damage actually does cause conscious deficits, this confirms HO theory.[8]

So far, so good. But consider now that it not just *any* conscious deficits that would confirm HO theory, but rather the *right kind*. Say, for instance, that we have reason to predict that severe damage to an integral area will cause severe deficits of consciousness; if so, if a potential integral area is severely damaged, and this causes only *minor* conscious deficits, this could not be considered to confirm HO theory. So, if we want to know whether any given lesion study confirms HO theory, what we need is a well-developed account of what kinds of conscious deficit would occur when an integral area is damaged. Such an account is provided in Section 3, with a focus on visual consciousness in particular. (This article will only be concerned with visual consciousness.) The other thing that we will need is the location of the potential integral areas, a matter to which we now turn.

2.2 | The prefrontal cortex and potential integral areas

In the issue of where the potential integral areas might be located, there has been much convergence, in that many agree that they are most likely in the prefrontal cortex (PFC), this being true of both HO theory's advocates (Kriegel, 2007, 2009; Lau, 2008a, 2008b; Lau & Rosenthal, 2011a; LeDoux & Brown, 2017; Brown, Lau & LeDoux, 2019; Lau & Brown, 2019; but see Gennaro, 2012, 2016), and its detractors (Block, 2007, 2009; Kozuch, 2014). The PFC, which encompasses the foremost parts of the brain, is responsible for carrying out more complex activities, such as reasoning, decision-making, and goal-directed activity (Fuster, 2002). Most interesting, for present purposes, is that the PFC carries out *metacognitive* functions (Metcalfe & Schwartz, 2015), some of which plausibly involve the HO representation of perceptual states (Kozuch, 2014). For instance, the PFC plays a central role in working memory (Baddeley, 2003), a function that involves selecting, and perhaps therefore also representing, the content of lower-order states (i.e., the content to be retained in working memory).[9] Some HO theorists have thought that the dorsolateral PFC (dlPFC) is an especially strong candidate for being a potential integral area (Kriegel, 2007, 2009; Lau, 2008a, 2008b; Lau & Rosenthal, 2011a), largely on the basis of studies purportedly showing correlations between visual consciousness and dlPFC activity (e.g., Lau & Passingham, 2006; Sahraie et al., 1997). However, whether such correlations obtain are controversial (Kozuch, 2014). Perhaps better evidence for dlPFC being a potential integral area comes from numerous studies indicating its involvement in metacognition of visual states (Rounis, Maniscalco, Rothwell, Passingham & Lau, 2010; Chiang, Lu, Hsieh, Chang & Yang, 2014; Fleming, Ryu, Golfinos & Blackmon, 2014; Miyamoto et al., 2017; for review, see Vaccaro & Fleming, 2018).[10] Overall, there is good reason to locate potential integral areas in the PFC, and in the dlPFC in particular.

While most HO theorists regard the PFC as crucial for consciousness, they differ on what precise role it plays. One difference concerns whether they take PFC content to be what appears in

[8]To say that datum D "confirms" theory, T is to say that observing D increases the probability of T being true.

[9]Something currently debated is whether the PFC actually stores the representations selected for working memory, or merely manages representations stored elsewhere (e.g., in sensory areas) (for review, see Xu, 2017). As we will see (next paragraph), this debate has consequences for how HO theory should be neuroscientifically construed.

[10]While there is no consensus yet as to what precise role dlPFC plays in metacognition, it is commonly thought to participate in a metacognitive network, one that might include the anterior PFC (Shekhar & Rahnev, 2018) and/or the anterior cingulate cortex (Bang & Fleming, 2018), among other areas (Vaccaro & Fleming, 2018).

consciousness. According to "upper-deck" views (Brown, 2015; Lau & Rosenthal, 2011b; LeDoux & Brown, 2017), activity in the PFC constitutes the content of consciousness. In these views, an experience as of greenness, for instance, occurs when and only when an HO state represents a lower-order (LO) state as having greenness as its content, with it making no difference phenomenologically if the LO state represents some other color (e.g., redness), or even if the LO state does not exist (Brown, 2015; Lau & Rosenthal, 2011a; Rosenthal, 2011). According to "lower-deck" views (Kriegel, 2007; Lau, 2008b; Lau & Brown, 2019), it is activity in sensory areas that typically constitutes the content of consciousness; more specifically, the content of consciousness is typically constituted by those LO states in sensory areas that have been successfully targeted by HO states in the PFC.[11,12] In these views, one experiences greenness when and only when an HO state targets an LO state that (actually) has greenness as its content (and if it were the case that the LO state had a different color as its content (e.g., redness), it is that color that would be consciously experienced). An important difference between these views is that, in the upper-deck version, some content's becoming conscious typically involves the PFC "duplicating" content from LO sensory states (since the content of the LO state will be embedded in the HO representation targeting it), whereas in the lower-deck view, the HO state merely "refers" to the LO state in some manner, this referral being sufficient for making the LO sensory state's content conscious.[13]

Returning to the main point, it seems that the PFC is a good candidate for containing potential integral areas, with dlPFC being an especially strong candidate. But, regardless of whether or not these claims are true, we take them as assumptions, since the goal of this article is not to determine whether the PFC contains potential integral areas, but rather to determine whether, *if* this is true, the lesion studies to which HO theorists appeal actually support their theory.

2.3 | Lesion studies used as evidence for HO theory

This subsection examines studies that HO theorists (e.g., Lau & Rosenthal, 2011a; Lew & Lau, 2017; Morales & Lau, 2020) have used to support their theory, also discussing why one might take them to do so. We start by describing an experimental paradigm used in both.

In *visual masking* (Breitmeyer & Ogmen, 2000), the subject is shown two stimuli (e.g., two letters) in quick succession, in either the same or adjoining parts of the visual field; the first stimulus is known as the "target" the second as the "mask." If the stimuli are presented in the right way, the subject reports not having seen the target; it is "masked." What is crucial to the stimulus being successfully masked is that the interval between the target and mask be not too large or small (around 60–100 ms is often effective, Bugmann & Taylor, 2005). Now we examine two of the studies to which HO theorists have appealed (more are presented in Section 4.1).

The first is due to Del Cul, Dehaene, Reyes, Bravo and Slachevsky (2009). In this experiment, the target was a letter, and the mask was four letters arranged in a cross. Both stimuli were presented at fixation, with the interval between presentations being varied. The

[11]I say "typically" because this could possibly occur in other areas; for instance, Lau and Rosenthal (2011b) hypothesize that cases where an HO state itself becomes conscious involves HO states in one part of the PFC being targeted by HO states in other parts of the PFC.

[12]In Lau and Brown's "joint determination" view, the HO state just determines the "intensity" with which the content is experienced.

[13]Picking up from fn. 10: It is worth pointing out here that it seems that upper-deck views can appeal to working memory as an "integral function" only if the PFC-storage theory of working memory is correct; conversely, it seems that lower-deck views can do so only if the sensory-storage theory is correct.

experimenters found that subjects with PFC damage, on average, required a longer interval between the target and mask than did normal subjects before reporting seeing the target (about 20 ms extra). It seems that normal subjects consciously experienced stimuli that PFC patients did not.

The second study is due to Rounis et al. (2010). In this experiment, the targets were a square and triangle presented next to each other, with them having an either left–right or right–left arrangement; the masks were two stars presented in the same parts of the visual field as the targets. In each trial, the subjects were asked to identify the arrangement of the target stimuli, also rating how "clear" they were, relative to other trials. In one condition, transcranial magnetic stimulation (TMS) was used to decrease cortical excitability in the dlPFC,[14] after which subjects on average rated the targets as being less clear. If one takes this lack of clarity to indicate a lack of consciousness, then it seems as if the TMS prevented conscious perception of the targets.

Let us refer to these studies, along with the other ones to which HO theorists appeal to support their theory, as the "PFC Studies" (the others are examined in Section 4.1). Now, when HO theorists appeal to the PFC studies to support their theory (Lau, 2011; Lau & Rosenthal, 2011a; Morales & Lau, 2020), they leave unsaid as to why the studies should be taken to do so. But we can suppose that it is because HO theory seems to provide a good explanation for these data, in that the deficits in consciousness can be thought to have resulted from the PFC lesions having caused a loss of HO states. For instance, in the Del Cul et al. study, the idea would be that the PFC patients fail to consciously represent the target because their lesions prevented the appropriate HO state from being formed, namely, the one that would have targeted, and therefore made conscious, the content of a lower-order representation of the target. Similarly, in the Rounis et al. study, the idea would be that in those cases where subjects report not consciously seeing the stimulus (i.e., when they rate it as "unclear," relative to other trials), this is because the TMS caused certain HO states to be lost, namely, those targeting lower-order representations of the square and triangle.

It seems, then, that there is an interpretation of these studies according to which they confirm HO theory.[15] However, as discussed above, whether these studies actually do so depend on what kinds of conscious deficit should occur when an integral area is impaired. This is the subject of the next section.

3 | NEUROPSYCHOLOGICAL PREDICTIONS OF HO THEORY

While the assumption that integral area damage should produce conscious deficits seems to be in use whenever lesion data are used to support or undermine HO theory, there has been little discussion as to what precisely these deficits should be like (but see Kozuch, 2014). This section investigates this issue, arguing that HO theory must predict integral area impairment to likely result in a large number of visually conscious states being lost. Now we go through the argument step by step.

[14]In this experiment, a particular kind of TMS was used, *theta burst* TMS (Huang, Edwards, Rounis, Bhatia, & Rothwell, 2005). In this technique, stronger pulses of magnetic stimulation are used to decrease cortical excitability for a relatively long time (in this experiment, up to 20 min). The TMS here causes neurons to act as if they have undergone long-term potentiation or depression, the result being that the brain area's ability to function is hampered; in the Rounis study, the application of TMS to dlPFC was shown to affect subjects' ability to produce accurate metacognitive judgments.

[15]As will see below (Section 4.1), the HO theorist might try to explain these data in another way, which is that the PFC damage does not cause HO states to be lost, but rather just to *misrepresent* the content of a lower-order state.

3.1 | Integral area impairment probably results in a deficit in HO states

When an integral area is impaired,[16] this would plausibly result in a reduction in the number (or complexity)[17] of HO states that the integral area could produce. One would naturally think that this leads to a significant deficit in the overall number of HO states that the lesioned subject could produce. It has been argued, however, that the deficit in HO states created by integral area impairment might be largely or completely made up for by other integral areas, or by brain areas enlisted as such, perhaps through neuroplasticity (Kriegel, 2007, 2009; Lau & Rosenthal, 2011b).[18] While cogent arguments have been given (Kozuch, 2014; cf., Kozuch, 2015) for the unlikelihood of neuroplasticity having brought about a complete (or near complete) recovery of function in the PFC Studies (see also Grafman, Zahn & Wassermann, 2010),[19] some unique properties that PFC neurons have (Mante, Sussillo, Shenoy & Newsome, 2013) make the issue worth revisiting, something that we do in Section 4.2. For the time being, the idea that the lesions in the PFC studies have resulted in a significant deficit in HO states will be taken as an assumption, perhaps with good reason: As seen above, the idea that the PFC studies support HO theory are *predicated* upon HO states having been lost (since these are supposed to explain the purported conscious deficits), and so, were there no lost HO states, the issue of whether these studies support HO theory would already be resolved—not in the HO theorist's favor.

3.2 | A deficit in HO states causes an absence of conscious visual content

We just saw that integral area impairment likely manifests as a reduction in the amount of HO states. Now we look at the issue of what the phenomenological results of such a reduction would be.

Approaching this issue requires first introducing a distinction between ways in which one's visual experience might change, this being between *differences* and *absences* of conscious content within one's visual experience. Consider a subject that is having an experience as of redness in central vision, whose experience of color then changes, in one of two ways: In the first case, the subject starts to experience some color other than red in central vision, say, experiencing green instead; this is a *difference* in content. In the second case, the subject starts to lack an experience as of color in central vision; this is an *absence* of content. The two scenarios differ importantly in that only an absence of content involves a change in the *amount* of conscious content: In the case involving a difference, the subject's experience as of redness in central vision was *replaced* by an experience as of greenness in central vision, so there was no loss of conscious content; in the case involving an absence, the subject went

[16]Here is what I mean by "integral area impairment": Some integral area A is impaired if and only if (a) A has been lesioned and (b) the lesion means that A can no longer produce HO states as effectively as it had before.

[17]The parenthetical clause here covers superficially different interpretations of HO theory, ones according to which it is not a large number of HO states creating our visual phenomenology, but rather just one much more complex HO state.

[18]The term "neuroplasticity" refers to the brain's ability to reorganize itself in response to injury or experience (Grafman, 2000). When neuroplasticity happens in response to a lesion, it is something usually occurring over long periods of time (Ibid.), though it can sometimes occur dynamically (Voytek et al., 2010).

[19]Putting it generally, Kozuch's point here is that such a scenario is unlikely to obtain because it either implies that the brain produces HO states in an implausibly redundant fashion, and/or assumes a form of neuroplasticity more powerful than exists in actuality.

from experiencing redness in central vision, to not experiencing any color in central vision, so there was a loss of conscious content.

The reason that the distinction between differences and absences of content is important is because—as I argue now—a deficit in HO states results in an absence, and not just a difference, in one's conscious visual content. Consider the following: According to HO theory, what makes a state conscious is its being targeted by an HO state. Given this, HO theory must say that, if a creature has conscious states at all, it is in virtue of the creature having produced one or more HO states, that is, the ones targeting the lower-order states. Imagine now a creature whose mental states have never been targeted by HO states; such a being (according to HO theory) is entirely unconscious, in the sense of there being nothing that it is like to be that creature; put another way, it has no conscious content. Picture it now becoming the case that one of our "zombie" creature's perceptual states comes to be targeted by an HO state; at this point, the creature would have a conscious state; for the first time, the creature now has conscious content. Say that the lower-order state being targeted is one representing redness. This means that now a little bit of conscious color content (viz., an experience as of redness) comes into existence. Compare this now to what happens if the HO state goes away: Now the experience as of redness goes out of existence; the creature reverts to no longer having conscious content.

I think that a case like this makes vivid the idea that losing an HO state would cause an *absence* of conscious content, and not merely a *difference*: When the creature loses its sole HO state, what happens is not that the experience as of redness changes to an experience as of another color (say, green); rather, the experience just ceases to exist, the conscious content going absent.

It seems, then, that there is good reason to think that deficits in HO states would cause absences of conscious content. This seems true, furthermore, in the case of both upper-deck and lower-deck versions of HO theory: In the upper-deck version, the loss of HO states means that whatever content these states possessed now does not show up in the subject's experience; in the lower-deck version, the loss of HO states means that the content of whatever LO states would have been targeted now does not show up in consciousness. The idea that losing HO states causes a loss in conscious content becomes important later, when we reanalyze the experiments to which HO theorists have appealed. Next, we investigate the issue of what such absences would be like.

3.3 | A loss in conscious visual content phenomenologically resembles blindness or visual agnosia

So far it has been established that integral area impairment likely results in a deficit of HO states, and that this would lead to an absence of conscious content. Now we investigate the issue of what the phenomenological results of such a loss would be, focusing on visual consciousness (cf., Kozuch, 2014). As it turns out, there are visual disorders that serve as examples of lost conscious visual content, ones that will eventually be used as models for what results when an integral area is impaired.

We start with instances where there is a *complete* loss of conscious visual content. The model for this would be complete blindness, this being a case in which a subject loses *all* of their conscious visual content (Symonds & MacKenzie, 1957). Such a case is simple to understand; more complicated are instances where just *part* of one's conscious visual content is lost.

Understanding what would happen in such cases involves first noting something about visual experience, which is that it largely consists of the representation of various properties (e.g., color, shape, motion), with each of these properties being represented as occurring in some part of the visual field (e.g., in

central vision, just to the right of central vision, just above central vision, etc.). Building off this, it seems that, in cases where a subject loses less than all conscious visual content, there are two ways that this might happen. The first would be that the subject loses, in certain parts of visual field, the ability to experience *any* visual properties. This would phenomenologically resemble *partial blindness*. Some examples of partial blindness would be *hemianopia*, blindness in half of one's visual field, or a *scotoma*, blindness occurring in some localized portion of the visual field. A second way in which one might lose conscious visual content is through losing the ability to represent some *certain kind* of visual property, where this deficit extends to all or just certain parts of the visual field (just as it can with blindness). Examples of these type-specific losses of conscious visual content come from the various *visual agnosias*. In *achromatopsia*, one loses the ability to experience object color, making those parts of the visual field that are affected phenomenologically appear monochromatic (e.g., black and white) (Damasio, Yamada, Damasio, Corbett & McKee, 1980; Zeki, 1990). In *akinetopsia*, one loses the ability to experience the motion of objects, causing the world to appear as if one is watching a movie with missing frames (Walsh, Ellison, Battelli & Cowey, 1998; Zihl, Von Cramon & Mai, 1983). Other agnosias include *visual form agnosia*, the inability to experience object shape (Heider, 2000), or *associative agnosia*, the inability to discern object identity (McCarthy & Warrington, 1986).[20]

The above provides some general models for what it is like to lose conscious visual content, something needed to construct an account of what occurs when an integral area is impaired. The next subsection produces this account.

3.4 | The phenomenological results of integral area impairment

Consider a case in which integral area impairment creates a deficit in HO states. Because some HO states previously being produced were no longer being produced, it would cause some of the conscious visual content that was formally a part of the subject's experience to go missing, with large or small "gaps" appearing in certain parts of the visual field (cf., Kozuch, 2014). The gaps here would be patches of the visual field in which the subject experiences something phenomenologically the same as one or more types of blindness or agnosia. I will explain.

Let us start by considering a case in which all integral areas are *destroyed*, that is, cases where there is a *complete* loss of ability to produce HO states. Since HO states are necessary for conscious states, this would result in a *complete* lack of conscious visual content; it would result in a condition phenomenologically the same as complete blindness (we can just call this "phenomenological blindness"). But cases where the ability to produce HO states is completely lost would probably occur rarely, if ever. More frequently, integral area impairment would only result in a *partial* loss in the ability to produce HO states. In such cases, the exact manner in which the conscious deficits manifest would be determined by which lower-order (LO) states now fail to be targeted. We look at toy examples.

Consider a case in which integral area impairment has caused tokens of some higher-order state H to be no longer produced, though they were reliably produced before the lesion; say, furthermore, that H's job had been to represent tokens of lower-order state L, where L's job was to represent object properties in central vision. It seems now that, since tokens of H are no longer formed, the subject would experience a loss of conscious visual content in central vision; they would experience a centrally located phenomenological scotoma. Consider now a different case, one where L's job is not to represent just

[20]These deficits can all be caused by damage to specific areas in the occipital (and sometimes temporal) cortex. For instance, complete and partial blindness result from primary visual cortex lesions, and achromatopsia or akinetopsia occurs when mid-level brain areas V4 or V5 are damaged, respectively.

any visual property in central vision, but specifically color. Here, the subject experiences a loss of conscious color content in central vision; they experience phenomenological achromatopsia there.

So far, we have examined simple examples, ones involving just one type of HO state missing, and thus just one type of LO state no longer becoming conscious. In practice, the phenomenological results of integral area impairment would almost certainly be much messier, the exact results hard to predict. Explaining why requires first making some general observations about conscious visual experience, and how it would relate to HO states.

3.4.1 | Constructing a typical visual experience requires numerous and varied HO states

Earlier, the observation was made that visual experience consists of many represented properties, where each property is represented as being in some certain parts of the visual field. Notice now that visual experience seems to consist of *a lot* of represented properties, represented in *many* parts of the visual field: At each point in the visual field, things like colors and textures of objects might be represented, often along with their shape, motion, distance, and (perhaps)[21] the category to which the object belongs. And so it seems that visual experience contains a lot of content, in that it seems to contain numerous conscious states, ones coming in a wide variety (see, e.g., Gregory, 1966; Siewert, 1998, Chapter 7; Carruthers, 2000, Chapter 8).

Now consider that, according to HO theory, for each of these varied and numerous conscious visual states, there must be a corresponding HO state, namely, the one targeting the appropriate LO state (Carruthers, 2000).[22] Given the amount of content in visual experience, this is a lot of HO states: In typical visual experience, for each point in the visual field, any of a number of properties (motion, color, shape, etc.) might be consciously represented there simultaneously, and for each of these properties, an HO state (or portion thereof)[23] must target whatever LO state represents this property. And so it seems that the thesis of this subsection is true: Constructing a typical visual experience would require numerous and varied HO states.[24]

[21]As discussed below (Section 4.2), it is controversial as to whether representations of the category to which an object belongs are ever *phenomenally* conscious.

[22]Carruthers writes, "According to ... [higher-order thought theory], I ... need to have a distinct ... higher-order belief for each distinct aspect of my experience—either that, or just a few such beliefs with immensely complex contents" (2000, p. 221). While this issue receives little discussion, most HO theorists seem to accept this entailment of the theory (e.g., Rosenthal, 2002).

[23]This covers the possibility of "complex" HO states discussed in fn. 18.

[24]Not long ago, experimental phenomena such as inattentional and change blindness were used to argue against the "richness" of visual experience (Blackmore, 2002; Blackmore, Brelstaff, Nelson, & Trościanko, 1995; Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Dennett, 1991; O'Regan, 1992; Rensink, 2000), and considerations raised by these researchers have been used to argue against visual experience requiring a large number of HO states (Gennaro, 2004; Weisberg, 1999). However, arguments advanced for visual experience being impoverished have not stood the test of time (Block, 2001; Cohen, 2002; Noë, Pessoa, & Thompson, 2000; Simons & Ambinder, 2005; Wolfe, 1999), and more recent attempts to argue this (Knotts, Odegaard, Lau, & Rosenthal, 2019) seem to establish just differences in conscious content, not absences. But even if these arguments succeeded in showing visual experience to not contain as much content as we thought it did—for instance, if it were shown that visual experience is detailed only in central vision (Dennett, 1991, Chapter 11)—visual experience would still probably contain a lot of content: Just in central vision, we find many, many points at which multiple properties such as color, shape, motion, distance, and/or object identity might be represented at any given moment (cf. Carruthers, 2000, pp. 299–301).

3.4.2 | The phenomenological results of integral area impairment

As discussed above, integral area impairment likely results in deficits in HO states. What exactly this would be like phenomenologically would be determined by the distribution of HO states that were lost. While it is difficult to predict exactly what this distribution would look like, there are three axes upon which the deficits could vary, allowing us to map the possibilities. These three axes concern the degree to which the deficits are prevalent, diffuse, and dynamic. I discuss each in turn.

A loss in conscious visual content caused by integral area impairment can be more or less *prevalent*, according to how many fewer conscious states the lesion caused the subject to have, relative to the amount usually composing their visual experience. The prevalence of the lost conscious states will be commensurate to the loss in HO states. For instance, if all integral areas are completely destroyed, meaning no HO states can be produced, this would result in the subject having no conscious visual content; it would cause complete phenomenological blindness. (The qualifier "phenomenological" is henceforth mostly left implied.) Correspondingly, if the integral area impairment means that only half as many HO states could now be produced, this would result in a visual experience containing half as much content; one way this could potentially (if improbably) manifest is if the subject were left only with conscious states representing properties in just half of their visual field, resulting in (phenomenological) hemianopia.

The second axis upon which conscious deficits might vary is the degree to which they are *diffuse* or *localized*. Consider a case in which the lost HO states are those whose job had been to target a specific set of LO states, these being LO states representing contiguous portions of central vision; this would create one centrally located scotoma, in which case we could say that the conscious deficits are *localized*. Consider a different case now, one where the lost HO states are those targeting LO states that represented non-neighboring parts of the visual field; this would result in multiple smaller scotoma throughout the visual field, in which case we could say that the conscious deficits are *diffuse*. So far, we have discussed diffuseness in a *spatial* sense, that is, how deficits can be diffuse or localized *in the visual field*. Also important is the idea that deficits can be diffuse in a *content* sense, that is, the deficits could be more or less diffuse among *types* of visual experience. For instance, it could be that the only HO states lost were those targeting LO states representing color, in which case the deficits are *localized* to color experience. Another possibility is that the LO states that are no longer being targeted are distributed among multiple types of visual content; for instance, it could be that they are distributed among color, motion, and shape representations, in which case the deficits are *diffuse* among color, motion, and shape experience (and consist of color, motion, and shape agnosia). There are, of course, *many* other ways in which the conscious visual deficits might manifest, given the large number of permutations created by the potential for these deficits to be diffuse in varying degrees, and across multiple types of visual content and places in the visual field.

The next axis upon which the conscious deficits could vary is the degree to which they are *stable* or *dynamic*. Consider a case in which the lack of an HO state *consistently* results in the same LO states failing to be targeted, where these LO states had the job of representing colors in central vision. In this case, it seems that the subject would have permanent color agnosia for central vision. Because the loss in conscious visual content is always in the same type of experience, we can say that it is *stable*. Now consider a case where the LO states failing to be targeted vary between LO states that represent either color or motion. Because the loss in conscious visual content changes in what type of LO state fails to be represented (sometimes color, sometimes motion), we can say that the loss of content is *dynamic*. A loss in conscious visual content

also has the potential to be dynamic in a *spatial* sense, with it varying as to what part of the visual field fails to be consciously represented. Like in the case of diffuseness, many permutations of dynamic deficits are possible, given that they can be dynamic to any degree, and among any types of visual content or places in the visual field.

Something seen in this subsection is that deficits in conscious visual states caused by integral area impairment could vary greatly in how they manifest, in that such deficits could vary along several dimensions, these being how prevalent, diffuse, and dynamic they are, and in whether these variations occur spacewise or contentwise. That HO theory would predict integral area impairment to potentially cause deficits as disparate as those just described is an interesting consequence of the theory, in part because this might be used to help disconfirm HO theory (see Section 4.3). Next, we see what general predictions can be made on the basis of our investigation so far.

3.5 | A neuropsychological prediction of HO theory

What we have learned so far is that integral area impairment would produce deficits in visual consciousness, where these can vary in how prevalent, diffuse, and dynamic they are. Now, given the high variability with which the deficits could appear, what can we say with confidence about the results of integral area impairment? While currently limited neuroscientific knowledge makes it difficult to say much with certainty, there is reason to think that such deficits would be diffuse and dynamic, and stronger reason to the think that they would be prevalent.

That we would expect the deficits to be diffuse and dynamic comes from some properties possessed by PFC neurons. One of these is *mixed selectivity*, which is the ability of a neuron to simultaneously participate in multiple representations (Fusi, Miller & Rigotti, 2016; Rigotti et al., 2013). Importantly, neurons with mixed selectivity tend to use a "denser" coding than do neurons in visual areas, which means that larger numbers of neurons are used to represent any given visual feature, with it varying over time as to which populations of neurons constitute a given representation (Parthasarathy et al., 2017).[25] Another property that PFC neurons possess is *non-retinotopicity*, which means that it is relatively infrequent that PFC neurons located adjacent to one another will represent adjacent parts of the visual field (Jerde & Curtis, 2013), this again being in contrast to neurons in visual areas (Wandell, Dumoulin & Brewer, 2007) .[26]

We would guess that the mixed selectivity of PFC neurons would mean that damage here would lead to deficits that are dynamic, since PFC neurons' ability to constitute different representations at different times makes it more likely that the neural resources remaining after a lesion would be rededicated moment to moment; this in turn would lead to high moment-to-moment variance as to which LO states fail to be represented, and therefore also in which visual properties and parts of the visual field fail to be consciously experienced. Additionally, both non-retinotopicity and mixed selectivity make it seem more likely that the absences of conscious states arising post-lesion would be diffuse around the visual field, rather than localized. Indeed, with less severe lesions, phenomenological *scotomas* might not often appear, since

[25]Notably, that PFC neurons are mixed selective and use dense coding might appear to provide reason to expect the PFC to be more resilient against lesions than other brain areas, an issue to which we return in Section 4.2.

[26]In the PFC, the spatial arrangement of the representations might instead be organized by the "priority" of certain parts of visual field, as determined by task-relevance (Jerde & Curtis, 2013).

mixed selectivity and non-retinotopicity might make it less likely that there would be a simultaneous loss of *all* consciously represented properties in some part of the visual field.

It seems, then, that there is some reason to expect conscious deficits caused by integral area impairment to be diffuse and dynamic. As discussed below (Section 4.3), the deficits would be of this nature helps provide a way to disconfirm HO theories. However, this prediction is of secondary importance in the case of what this article is focused on evaluating, that is, the idea that the PFC studies confirm HO theory. What is of primary importance is that there is reason to think that deficits created by integral area impairment would be of high *prevalence*, in that we should expect it to typically cause the loss of many visually conscious states. I will explain.

It seems certain that, when an integral area is heavily impaired, there would be a large number of conscious states lost. For instance, if the impairment made it so that only half as many HO states could be formed as before, the subject would be bereft of half of their visually conscious states; this of course is a scenario in which a large number of conscious states has been lost. But many conscious states would be lost even in instances of mild impairment. Consider a case where the lesion resulted in only nine-tenths of the amount of HO states being able to be produced; while it would be only one-tenth of the conscious states that were lost, this would be one-tenth of a very large number of conscious states; more precisely, it would be one-tenth of all those conscious states necessary for populating the detailed visual experience that we have, one in which properties such as color, motion, and shape are represented at many different points in the visual field. And so it seems that, in cases where an integral area is significantly impaired—be this impairment light or heavy—many conscious states are likely to be lost.

If we combine this conclusion with the idea that the PFC is probably the best candidate for containing integral areas, we get a prediction of HO theory, which is that damage to areas in the PFC will sometimes produce the loss of many visually conscious states. Given this, we can say that, if a study provides evidence for PFC damage having caused the loss of many visually conscious states, then that study confirms HO theory. Consider though, that we can go further on the basis of the investigation carried out above, making the appearance of such deficits a *necessary* condition on a PFC lesion study confirming HO theory; that is to say, since the considerations presented above make it unlikely that integral area impairment would result in the loss of fewer than many visually conscious states, we can take the following thesis to be true:

> A PFC lesion study confirms HO theory only if the lesion causes the loss of many visually conscious states

This thesis in hand, we now return to the studies to which HO theorists have appealed to support theory (the "PFC Studies"), so as to evaluate whether they actually do.

4 | THE PFC STUDIES DO NOT CONFIRM HO THEORY

In Section 2, we examined the "PFC studies," a collection of lesion experiments sometimes thought to support HO theory, on grounds that the PFC damage seems to have produced deficits in visual consciousness. In Section 3, we investigated the issue of what kinds of conscious deficits should result from integral area impairment, arriving at a principle that can be used to judge whether a lesion study supports HO theory. Now this principle is used to argue that the PFC studies do not confirm HO theory.

Before proceeding, I should make clear that the goal of this section is not to establish a similar but stronger conclusion, this being that the PFC studies *dis*confirm HO theory. While this plausible idea is discussed further in Section 4.3 (and argued for in Kozuch, 2014), establishing

it would go beyond available space, and so this section focuses only on the less strong conclusion, that is, that the PFC studies fail to disconfirm HO theory.

Now we look at the PFC lesion studies to which HO theorists have appealed, along with some to which they might appeal in the future. As we will see, each study fails to demonstrate the loss of many visually conscious states, in which case none of them confirm HO theory.

4.1 | Reevaluating the experiments to which HO theorists appeal

We start with the Del Cul et al. study. In this experiment, PFC patients were more susceptible to visual masking than were controls, in that they would not report seeing the target until there was a larger interval between the target (a digit) and the mask (four digits arranged in a cross). As I explain now, this study does not confirm HO theory, since it does not provide evidence for the lesions having caused the subjects to lose any more than a very small number of conscious states.[27]

We start by considering those parts of the visual field outside of where the target is located. Here, there is no evidence for deficits of color or distance content; that is, the experiment's results provide no evidence for the lesioned subjects failing to enjoy normal visual experience in these parts of the visual field, in that the subjects apparently continue to experience the color and distance of whatever surfaces appear there (e.g., the stimulus background), just as they did pre-lesion. There is also no evidence for deficits in shape experience, since the subjects apparently continue to experience, where the mask is located, the shape of the numbers composing it. Similarly, there is no evidence for them not experiencing the category of the digits composing the mask—that is, there is no evidence for a lack of what is sometimes called "categorial" content—since the subjects seem to continue to experience the identity of the numbers composing the mask.

Moving on now to the part of the visual field where the mask is located: Here, the HO theorist might claim that there is a loss of color, shape, distance, and/or categorial content, on basis of subjects having reported not seeing the target, since one might take their having reported not seeing it as evidence for them not consciously representing these properties of the target. It is unclear, however, whether even this inference is safe, since what is being observed here could be just a failure to *report* on the stimulus, rather than a failure to consciously experience it (Block, 2005; Kozuch, 2014; Prinz, 2000; Zeki, 2003).[28] But I put this complication aside (though it is attendant to each PFC study), assuming that the target is not consciously experienced. Even so, the data still do not count toward a confirmation of HO theory, at least in the case of color and distance content. This is because what is required for confirming HO theory is that there be evidence that an integral area has been impaired, but instances of integral area impairment are expected to create *absences* in conscious content, and not merely *differences* (see Section 3.2); and in the case of color and distance content, there are merely differences. To see why, consider that, while the evidence might support the lesioned subjects failing to experience the *target's* color, it does not imply a *lack* of color experience where the target is located: If one examines their experience when viewing masked stimuli, it does not phenomenologically resemble some

[27]In the following discussion, I leave out one type of experience discussed above (in Section 3.4.1), motion experience, since the experiments do not involve moving stimuli.

[28]According to these commentators, the failure to report on the target plausibly occurs, not from a loss of conscious states, but rather a loss in the ability to produce reports about hard-to-see stimuli like the target in this experiment.

kind of color agnosia (or, for that matter, a scotoma), one where the deficit is localized to the target's location; rather, one seems to experience the color of the background.[29] Since the subject experiences the background's color instead of the target's, this is a *difference* in content, not an absence, in which case there is no conscious state lost. The same can be observed about *distance*, since one does not appear to lack distance content where the target appears, but rather consciously perceives the distance at which the background is located.[30]

Lastly, we consider shape and categorial content where the target is located. Here it seems that we might finally have evidence for an absence of conscious content: Subjects report not seeing the identity of the stimulus, and this seems to indicate a lack of categorial experience where the target is located, the same being true of shape experience (given that the target's shape is probably what subjects use to consciously identify what digit it is). However, while this might be evidence for lost conscious states, it is only a small amount, namely, just those states representing the target's shape and category.

Overall, the Del Cul study does not provide what is needed to confirm HO theory: Of the numerous and variegated conscious visual states typically composing visual experience, the only ones for which there is evidence that they have been lost are those representing properties in the part of the visual field where the target appears, and only a small portion of these, namely, category and shape (and just *possibly* these; see just above, and fn. 29). On the other hand, there is no evidence for deficits of color or distance content in any part of the visual field, or in shape or categorial content in parts of the visual field outside the target's location. Since this study presents evidence for only two visually conscious states having been lost, and what would be needed for confirming HO theory is many visually conscious states having been lost, the Del Cul study does not confirm HO theory.

The same is true of the other experiment examined above, the Rounis et al. metacognition study. The reader will remember that subjects in this experiment were shown masked stimuli (a square and triangle, each masked by a star), their task being to say whether the square or triangle appeared on the left. The relevant result in this experiment was that, after TMS was used to depress activity in the dlPFC, subjects rated the targets as being "less clear" (i.e., relative to the targets' perceived clarity on other trials). Something important about this experiment is that the masks were presented in the same place as the targets. This means that there was no loss of color, distance, shape, or category content in this part of the visual field, even when the targets were successfully masked: Even if the masking caused the subjects to fail to experience these properties of the targets, the subjects *would* experience the color, distance, shape, and category of the stars masking them; this means that there is just a difference in content, not an absence. Additionally, and just like in the Del Cul study, there is no evidence for the lesion having caused conscious color or distance deficits anywhere in the visual field. Overall, the Rounis study seems to not provide evidence for *any* conscious states having been lost; it therefore also fails to confirm HO theory.

[29]A visual masking demonstration is found in the links below. Notice that in the instance where the target (a small black disk) is masked, what one experiences where the target is located is not an absence of color experience, but rather the background's color. (Each of the demonstrations starts with a cross that is to be fixated on, then the listed stimulus follows shortly after.)

1: Just the target

2: Target and mask

3: Just the mask

These demonstrations, acquired from this website, are used with permission of their owners (psychologists John Krantz and Bennett Schwartz).

[30]The demonstration in the last footnote supports this claim.

The other studies to which HO theorists appeal also do not confirm HO theory: In (Barcelo, Suwazono & Knight, 2000), subjects were presented with a rapid stream of triangles in which an inverted triangle randomly appeared, the subjects' task being to accurately detect its presence as quickly as possible. Subjects with unilateral dlPFC lesions were less accurate than controls when the stream of triangles was presented in the visual field contralateral to the lesion. This study presents no evidence for lost conscious content: Similar to the Rounis study, it seems that even if the target's properties were not consciously perceived, the properties of one of the distracting stimuli (i.e., an upright triangle) would be. Another experiment claimed to support HO theory is (Turatto, Sandrini & Miniussi, 2004). In this study, four faces arranged in a square were shown briefly to the subject, then shown again, with one face being changed on some trials. The subject's task was to say whether any faces had changed. It was shown that subjects that had TMS applied to right dlPFC were significantly less accurate in identifying whether a change had occurred than those given either left dlPFC or sham TMS.[31] Like the previous studies, the results here provide no evidence for the lesion having caused absences of conscious content (e.g., color and distance) in parts of the visual field outside where the target was located. But this is also true where the target *is* located: Instances of "change blindness" (Simons & Levin, 1997) like the one occurring in this study are most plausibly interpreted as cases in which the subject does not fail to experience things like the shape and color of the target (both before and after the change), but rather just the *change* in the target (Kozuch, 2019), this being an interpretation now accepted even by advocates of attention being necessary for consciousness (De Brigard & Prinz, 2010; Prinz, 2012, Chapter 3). And so it seems that the only absence of conscious visual content that might be found in this study is in *change* content (if this even is a type of *phenomenally* conscious content),[32] and only at the target's location.

There are, in addition, other PFC lesion studies that some commentators (some of whom are HO theorists) have used to try to support the idea that the PFC is essential for consciousness (Odegaard et al., 2017), and which one might therefore hope to use to support HO theory. It is not clear, however, whether any of the following deficits resulting from PFC lesions can be interpreted as cases in which any (let along many) conscious states are lost: Being slower and less accurate when asked to report the moment that a rapidly presented stream of cars turns into a stream of faces (Philiastides, Auksztulewicz, Heekeren & Blankenburg, 2011); being less accurate when saying whether a face or scene is the same one as presented before (Lee & D'Esposito, 2012); and being worse at judging how confident they (i.e., the subjects themselves) should be when choosing which of two circles contains more dots (Fleming et al., 2014). Based on what we saw above, it is easy to see how, in each of these cases, the natural interpretation would be one in which there is no substantial (or simply no) loss in conscious visual content; the same seems true of other lesion studies to which PFC theorists have appealed (Chiang et al., 2014; Komura, Nikkuni, Hirashima, Uetake & Miyamoto, 2013).[33]

[31]That the increased susceptibility to change blindness would be brought about in the right dlPFC condition is thought to be caused by the special role that it plays in change detection (Beck, Rees, Frith, & Lavie, 2001).

[32]Similar to the case of categorial content (see Section 4.2), it is open to debate whether there is ever anything that "it is like" to bear a representation of something having changed.

[33]Another case to which HO theorists might wish to appeal is (Quraishi, Benjamin, Spencer, Blumenfeld, & Alkawadri, 2017), a study in which intracranial stimulation of lateral PFC caused a subject to become unresponsive, and sit there with a "dazed, fixed look." However, a subject's being unresponsive does not indicate whether or not that subject continues to enjoy a phenomenology (see discussion of akinetic mutism, and Robert Knight's unresponsive subject, in Kozuch, 2014). Additionally, the kind of intense, invasive stimulation of the PFC used in this experiment might plausibly also disrupt activity in connected brain areas, including those areas that are alternative candidates for being the neural basis of consciousness (e.g., sensory areas).

4.2 | The PFC studies do not confirm HO theory

Remember the principle arrived at in the last section, this being that a PFC lesion study confirms HO theory only if it leads to many visually conscious states being lost. The investigation just conducted shows that, in each of these studies, there is only evidence for the lesion having caused the loss of between zero and two conscious states. And so it seems that the studies to which HO theorists have appealed to support their theory fail to do so.

Perhaps, however, these studies are even further from confirming HO theory than we might have thought: Consider that, in all but the Del Cul study, any visually conscious states that were lost were just those composing *categorial* experience. But it is controversial as to whether there even is *such a thing* as categorial experience, in that notable arguments have been advanced for the idea that there is never anything that "it is like" to bear a categorial representation (Dretske, 1995; Clark, 2000; Tye, 2002; Prinz, 2012; but see Siewert, 1998; Siegel, 2006). So, of even the handful of lost "conscious" states observed in these studies, some might fall outside the explanandum of HO theory.[34]

One might wonder whether the HO theorist could bolster the case for the PFC studies confirming HO theory by appealing to the mixed selectivity and dense coding of PFC neurons (see Section 3.5). The idea here would that, because of these properties of PFC neurons, when damage prevents one neural population from representing an LO state, some other population takes on the role, preventing the loss of a conscious state. This response could appear especially promising, given that mixed selectivity and dense coding are thought to greatly increase the representational capacity of neurons, perhaps giving the PFC the ability to represent far more visual properties than it was typically called on to represent pre-lesion. Were the PFC resilient against damage in this way, there would be reason to think that integral area damage might sometimes lead to just very minor conscious deficits.

A first thing to note about this approach is its limited application: It cannot help the HO theorist in cases where no visually conscious states are lost (i.e., a majority of the PFC studies), since the idea that a PFC study confirms HO theory relies on a correlation having been established between conscious deficits and PFC lesions (see Section 3.1). But even in the case of those studies where one or more conscious states might have been lost, this line of response is not all that helpful. Consider that, while mixed selectivity provides theoretical reason to expect the PFC to be resilient, so far all of the empirical data only seem to speak against this idea: On the one hand, data recently offered in favor of this idea (Lau & Rosenthal, 2011a; Lew & Lau, 2017; Morales & Lau, 2020) fail to be compelling[35]; on the other hand, not only is it the case that complete or near complete recovery from brain damage is rare, especially in the case of adults (Grafman et al., 2010), there is also a wealth of data in which significant and long-lasting deficits are shown to follow from PFC lesions, including ones to the dlPFC (for review, see Szczepanski & Knight, 2014). Indeed, there is reason to think that, in the PFC studies

[34]To the extent that it is meant to explain conscious *experience*; see fn.2.

[35]Though I lack space for a study-by-study explanation as to why this is the case, it can still be pointed out that (a) the oft-cited Voytek et al. (2010) study fails to support PFC resilience because it does not demonstrate patients to show a return to a near-normal level of function, as would be needed to prevent losing many conscious states (see detailed discussion in Kozuch, 2014), (b) that one cannot take the Mackey et al. study (Mackey, Devinsky, Doyle, Meager, & Curtis, 2016) to establish PFC resilience without first rebutting the compelling arguments (presented by the study's authors themselves) for the lack of deficits from dlPFC damage not being attributable to neuroplasticity, and (c) in the case of the other studies to which Morales and Lau appeal, a cursory examination of them reveal that more would need to be said as to why they support PFC resilience before they could be taken as doing so.

themselves, the lesions have caused significant deficits: In the Fleming et al. experiment, for instance, the lesioned patients showed significant impairments in their ability to metacognitively estimate their task performance; and in the Rounis et al. experiment, the particularly strong kind of brain stimulation employed ("theta-burst" TMS) makes the occurrence of cognitive deficits especially likely. Overall, given the seeming improbability of PFC lesions having been made up for by mixed selectivity (or any other purported source of PFC resilience), it seems that other explanations of the apparent conscious deficits in these studies should be preferred (Prinz, 2000; Zeki, 2003; Block, 2005; Pollen, 2008; Kozuch, 2014).

Summing up: In the case of some of the PFC studies, they fail to confirm HO theory because no conscious states appear to have been lost; in the case of the other PFC studies, they can be taken to confirm HO theory only by making improbable assumptions about the plasticity of the PFC. Overall, the PFC studies seem to fall far short of confirming HO theory.[36]

4.3 | Do the PFC studies disconfirm HO theory?

While it has just been established that the PFC studies fail to provide evidence for HO theory, one might wonder whether a stronger conclusion could be drawn, which is that these studies actually act as evidence *against* HO theory. In some ways, this is quite plausible, given that the PFC is often thought to be the best candidate for containing potential integral areas, and that the studies reviewed above consist of instances in which PFC lesions fail to bring about the predicted loss in visually conscious states. There are, however, shortcomings to basing such an argument on just these particular studies: For one thing, we cannot be sure that the studies collectively present instances of each potential integral area in the PFC being lesioned, something perhaps necessary for considering these studies to count strongly against HO theory. Additionally, more would probably need be said about the possibility of PFC resilience (discussed in the last subsection) before we could establish with certainty the idea that the PFC studies disconfirm HO theory. And so this article has aimed to establish only a more modest claim, which is that the PFC studies to which HO theorists appeal fail to support HO theory.

Perhaps what is more interesting, however, is how the apparatus developed here could strengthen existing (and perhaps future) lesion-based arguments against HO theory. For instance, Kozuch (2014) has argued that available PFC lesion data speak strongly against HO theory, since they show there to be no area in the PFC such that, when it is damaged, this causes the kinds of conscious deficit to be expected when an integral area is impaired. While the argument Kozuch presents is taken by some as supporting the ideas that the PFC is not

[36]There is a move available here to the upper-deck (but not lower-deck) HO theorist, this being to argue that the *differences* in conscious content that are seen in some PFC Studies should count toward confirming HO theory. The idea here would be that the PFC lesion caused an integral area to malfunction, but that this did not result in an *absent* HO state, but rather a *misrepresenting* HO state. For instance, when the mask's color is experienced rather than the target's color, this would be explained as resulting from an HO state misrepresenting the content of an LO state, such that the HO state represents the LO state as having the mask's color as its content, though the LO state actually has the target's color as its content. This idea, however, is incongruent with how visual masking is thought to arise, since common explanations of it (Bugmann & Taylor, 2005) hypothesize LO representations of the target to be either absent, or just weaker than a simultaneously present representation of the mask. Given this, it is hard to see how it could be plausibly claimed that the HO state is misrepresenting; this in turn means that we lack evidence for an integral area malfunctioning; and this in turn means that we lack those correlations between PFC damage and conscious deficits needed to confirm HO theory.

essential for consciousness (Boly et al., 2017; Gennaro, 2016) and/or that HO theory is incorrect (Marvan & Polák, 2017), something preventing Kozuch's argument from being as strong as it could be is that there was, as of that time, no well-developed account of the phenomenological results of integral area impairment; this in turn made it harder to say whether Kozuch was correct when claiming that the conscious deficits created by integral area impairment would (at least in some cases) be so phenomenologically "dramatic" and "striking" that they would likely be detected in the course of studying subjects with an impaired integral area. But reflect now on the fact that, in the account provided in this article, the conscious deficits resulting from integral area impairment would phenomenologically resemble blindness and/or one or more types of visual agnosia and would have the potential to be diffuse or dynamic, both in where they appear in the visual field, and among types of visual experience. It seems that deficits like these would probably be dramatic and striking even in those (probably rare) cases in which they are relatively stable and circumscribed (e.g., if they consisted of a persistent lack of visual experience in some moderately sized part of the visual field). However, considerations discussed above (Section 3.5) about how such deficits would probably tend to be diffuse, dynamic, and prevalent make it seem likely that the phenomenological effects of integral area impairment would be radical and unstable: For instance, the resulting deficits could cause the subject to, in one moment, lack color in certain parts of the visual field, while having no visual experience in some other part; in the next moment, have a large centrally located phenomenological scotoma with normal conscious vision elsewhere; in the next moment, lack experience of one object's shape and distance while failing to experience a different object's color; and so on—it seems that such deficits would be dramatic indeed![37] Given all this, one would guess that the apparatus created in this article has the potential to—at least once developed and supplemented—significantly strengthen the idea that available PFC lesion evidence disconfirms HO theory.

In addition, the account of integral area impairment provided in this article might be used profitably in neuroscientific investigations of HO theory likely to be carried out in the future: Kozuch's argument against HO theory does not include a survey of lesion data outside of the PFC, leaving open the possibility that there are integral areas located there. It seems, then, that one way to further the debate over HO theory would be to carry out a survey of non-PFC lesion data, seeing if damage to any potential integral areas outside of the PFC produces the predicted deficits in visual consciousness. An account such as the one developed in this article would of course be of great use in such a survey, since the account supplies a better delineated criterion by which to determine whether the results of lesions to one of these areas confirm or disconfirm the idea that it is an integral area.

In summary, while what has been presented in this article probably cannot itself act as strong disconfirmation of HO theory, the ideas developed here might be used to greatly strengthen existing or future lesion-based arguments against HO theory.

[37]An epistemic issue lurks here, one perhaps threatening the possibility of using lesion evidence to confirm *or* disconfirm HO theory. The idea here would be that integral area impairment might cause not only deficits in consciousness, but also in introspection, making it difficult for the lesioned subject to report the dramatic changes in experience. This thorny issue—one that is not corrosive to the main point of this article (i.e., that the PFC Studies *do not confirm* HO theory)—must be left for future research. However, it seems unlikely that PFC lesions would typically *completely* disable a subject's introspective abilities, and it is probably also the case that a partial ability to introspect is good enough to recognize dramatic changes in one's experience; furthermore, it seems particularly likely that adequate introspective abilities would remain post-lesion if there turn out to be multiple introspective mechanisms (Hill, 2011; Prinz, 2004; Schwitzgebel, 2012).

5 | CONCLUSION

HO theorists have sometimes claimed there to be a set of PFC lesion studies that support HO theory, since they seem to present instances in which such lesions have led to deficits in consciousness. However, whether this is the case depends on precisely what kinds of conscious deficit should follow when an HO state-producing area is damaged. In this article, I argued that HO theory must predict such lesions to produce the loss of many conscious visual states, but that no such loss in conscious states occurs in any of these studies, meaning that these studies do not confirm HO theory. In addition, the ideas developed in this article have the potential to greatly strengthen both past and future lesion-based arguments against HO theory.

ACKNOWLEDGEMENTS

For insightful discussion and helpful comments, the authors would like to thank Chase Wrenn, along with those who attended presentations of this article that I gave at the Jean Nicod Institute and University of Alabama talk series. I am especially grateful to the two anonymous referees from Mind and Language for the valuable feedback that they provided. This work was made possible through grants from the University of Alabama Research Grants Council and the College Academy of Research, Scholarship, and Creative Activity.

ORCID

Benjamin Kozuch https://orcid.org/0000-0002-1550-213X

REFERENCES

Armstrong, D. (1968). *A materialist theory of the mind*. London: Routledge.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839.

Bang, D. & Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *115*(23), 6082–6087.

Barcelo, F., Suwazono, S. & Knight, R. T. (2000). Prefrontal modulation of visual processing in humans. *Nature Neuroscience*, *3*(4), 399–403.

Beck, D. M., Rees, G., Frith, C. D. & Lavie, N. (2001). Neural correlates of change detection and change blindness. *Nature Neuroscience*, *4*(6), 645–650.

Blackmore, S. (2002). There is no stream of consciousness. *Journal of Consciousness Studies*, *9*(5–6), 17–28.

Blackmore, S. J., Brelstaff, G., Nelson, K. & Trościanko, T. (1995). Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception*, *24*(9), 1075–1081.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247.

Block, N. (2001). Paradox and cross purposes in recent work on consciousness. *Cognition*, *79*(1–2), 197–219.

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, *9*(2), 46–52.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*(5–6), 481–499.

Block, N. (2009). Comparing the major theories of consciousness. In M. Gazzaniga (Ed.), *The cognitive neurosciences IV*. Cambridge, MA: MIT Press.

Block, N. (2019). Empirical science meets higher order views of consciousness: Reply to Hakwan Lau and Richard Brown. In A. Pautz & D. Stoljar (Eds.), *Blockheads! Essays on Ned Block's philosophy of mind and consciousness* (pp. 199–213). Cambridge, MA: MIT Press.

Boly, M., Massimini, M., Tsuchiya, N., Postle, B., Koch, C. & Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *Journal of Neuroscience*, *37*(40), 9603–9613.

Breitmeyer, B. G. & Ogmen, H. (2000). Recent models and findings in visual backward masking, a comparison, review, and update. *Perception and Psychophysics*, *62*(8), 1572–1595.

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

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

Bugmann, G. & Taylor, J. G. (2005). A model of visual backward masking. *Biosystems*, *79*(1–3), 151–158.

Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge: Cambridge University Press.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Chiang, T.-C., Lu, R.-B., Hsieh, S., Chang, Y.-H. & Yang, Y.-K. (2014). Stimulation in the dorsolateral prefrontal cortex changes subjective evaluation of percepts. *PLoS One*, *9*(9), e106943.

Clark, A. (2000). *A theory of sentience*. Oxford: Clarendon Press.

Cohen, J. (2002). The grand grand illusion illusion. *Journal of Consciousness Studies*, *9*(5/6), 141–157.

Damasio, A., Yamada, T., Damasio, H., Corbett, J. & McKee, J. (1980). Central achromatopsia behavioral, anatomic, and physiologic aspects. *Neurology*, *30*(10), 1064–1064, 1071.

De Brigard, F. & Prinz, J. (2010). Attention and consciousness. *Wiley Interdisciplinary Reviews, Cognitive Science*, *1*(1), 51–59.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J. & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E. & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*, *132*(9), 2531–2540.

Dennett, D. (1991). *Consciousness explained*. New York, NY: Little Brown & Co.

Dretske, F. I. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.

Fleming, S. M., Ryu, J., Golfinos, J. G. & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, *137*(10), 2811–2822.

Fusi, S., Miller, E. K. & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, *37*, 66–74.

Fuster, J. M. (2002). Frontal lobe and cognitive development. *Journal of Neurocytology*, *31*(3–5), 373–385.

Gennaro, R. (2004). Higher-order thoughts, animal consciousness, and misrepresentation: A reply to Carruthers and Levine. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness*. Amsterdam and Philadelphia: John Benjamins.

Gennaro, R. (2012). The consciousness paradox. In *Consciousness, concepts, and higher-order thoughts*. Cambridge, MA: MIT Press.

Gennaro, R. (2016). Higher-order thoughts, neural realization, and the metaphysics of consciousness. In P. Satsangi, S. Hameroff, V. Sahni & P. Dua (Eds.), *Consciousness: Integrating eastern and Western perspectives*. India: New Age Publishers.

Gennaro, R. J. (1996). *Consciousness and self-consciousness: A defense of the higher-order thought theory of consciousness*. Philadelphia, PA: John Benjamins Publishing Co.

Grafman, J., Zahn, R. & Wassermann, E. (2010). Brain damage: Functional reorganization. In *Encyclopedia of neuroscience* (pp. 327–331). Oxford: Elsevier Ltd.

Gregory, R. L. (1966). *Eye and brain: The psychology of seeing*. New York, NY: McGraw Hill.

Heider, B. (2000). Visual form agnosia: Neural mechanisms and anatomical foundations. *Neurocase*, *6*(1), 1–12.

Hill, C. (2011). How to study introspection. *Journal of Consciousness Studies*, *18*(1), 21.

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P. & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206.

Jerde, T. A. & Curtis, C. E. (2013). Maps of space in human frontoparietal cortex. *Journal of Physiology*, *107*(6), 510–516.

Knotts, J. D., Odegaard, B., Lau, H. & Rosenthal, D. (2019). Subjective inflation: phenomenology's get-rich-quick scheme. *Current Opinion in Psychology*, *29*, 49–55.

Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T. & Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nature Neuroscience*, *16*(6), 749.

Kozuch, B. (2014). Prefrontal lesion evidence against higher-order theories of consciousness. *Philosophical Studies*, *167*(3), 721–746.

Kozuch, B. (2015). Dislocation, not dissociation, the neuroanatomical argument against visual experience driving motor action. *Mind and Language*, *30*(5), 572–602.

Kozuch, B. (2019). Gorillas in the missed (but not the unseen): Reevaluating the evidence for attention being necessary for consciousness. *Mind and Language*, *34*(3), 299–316.

Kriegel, U. (2007). A cross-order integration hypothesis for the neural correlate of consciousness. *Consciousness and Cognition*, *16*(4), 897–912.

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

Lau, H. (2008a). Are we studying consciousness yet? *Frontiers of Consciousness: Chichele Lectures*, *2008*, 245.

Lau, H. (2008b). A higher order Bayesian decision theory of consciousness. *Progress in Brain Research*, *168*, 35–48.

Lau, H. (2011). Theoretical motivations for investigating the neural correlates of consciousness. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(1), 1–7.

Lau, H. & Brown, R. (2019). The emperor's new phenomenology? The empirical case for conscious experiences without first-order representations. In A. Pautz & D. Stoljar (Eds.), *Blockheads! Essays on Ned Block's philosophy of mind and consciousness* (pp. 199–213). Cambridge, MA: MIT Press.

Lau, H. & Passingham, R. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences*, *103*(49), 18763–18768.

Lau, H. & Rosenthal, D. (2011a). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373.

Lau, H. & Rosenthal, D. (2011b). The higher-order view does not require consciously self-directed introspection: Response to Malach. *Trends in Cognitive Sciences*, *15*(11), 508–509.

LeDoux, J. E. & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences*, *114*(10), E2016–E2025.

Lee, T. G. & D'Esposito, M. (2012). The dynamic nature of top-down signals originating from prefrontal cortex: A combined fMRI–TMS study. *Journal of Neuroscience*, *32*(44), 15458–15466.

Lew, S. & Lau, H. (2017). Crucial role of the prefrontal cortex in conscious perception. In *Executive functions in health and disease* (pp. 129–141). New York, NY: Academic Press.

Locke, J. (1690). *An essay concerning human understanding*. (A number of different editions of this book are available.)

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

Lycan, W. (2001). A simple argument for a higher-order representation theory of consciousness. *Analysis*, *61* (269), 3–4.

Mackey, W. E., Devinsky, O., Doyle, W. K., Meager, M. R. & Curtis, C. E. (2016). Human dorsolateral prefrontal cortex is not necessary for spatial working memory. *Journal of Neuroscience*, *36*(10), 2847–2856.

Mante, V., Sussillo, D., Shenoy, K. & Newsome, W. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84.

Marvan, T. & Polák, M. (2017). Unitary and dual models of phenomenal consciousness. *Consciousness and Cognition.*, *56*, 1–12.

McCarthy, R. & Warrington, E. (1986). Visual associative agnosia: a clinico-anatomical study of a single case. *Journal of Neurology, Neurosurgery and Psychiatry*, *49*(11), 1233–1240.

Metcalfe, J. & Schwartz, B. L. (2015). The ghost in the machine: Self- reflective consciousness and the neuroscience of metacognition. In J. Dunlosky & S. Tauber (Eds.), *The Oxford handbook of metacognition* (pp. 407–424). Oxford: Oxford University Press.

Miyamoto, K., Osada, T., Setsuie, R., Takeda, M., Tamura, K., Adachi, Y. & Miyashita, Y. (2017). Causal neural network of metamemory for retrospection in primates. *Science*, *355*(6321), 188–193.

Morales, J. & Lau, H. (2020). The neural correlates of consciousness. In *Oxford handbook of the philosophy of consciousness*. Oxford: Oxford University Press.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*(4), 435–450.

Noë, A., Pessoa, L. & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition*, *7*(1–3), 93–106.

O'Regan, J. K. (1992). Solving the real mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *46*(3), 461–488.

Odegaard, B., Knight, R. T. & Lau, H. (2017). Should a few null findings falsify prefrontal theories of conscious perception? *Journal of Neuroscience*, *37*(40), 9593–9602.

Parthasarathy, A., Herikstad, R., Bong, J. H., Medina, F. S., Libedinsky, C. & Yen, S.-C. (2017). Mixed selectivity morphs population codes in prefrontal cortex. *Nature Neuroscience*, *20*(12), 1770–1779.

Philiastides, M. G., Auksztulewicz, R., Heekeren, H. R. & Blankenburg, F. (2011). Causal role of dorsolateral prefrontal cortex in human perceptual decision making. *Current Biology*, *21*(11), 980–983.

Pollen, D. A. (2008). Fundamental requirements for primary visual perception. *Cerebral Cortex*, *18*(9), 1991–1998.

Prinz, J. (2000). A neurofunctional theory of visual consciousness. *Consciousness and Cognition*, *9*(2), 243–259.

Prinz, J. (2004). The fractionation of introspection. *Journal of Consciousness Studies*, *11*(7–8), 40–57.

Prinz, J. (2012). *The conscious brain*. Oxford: Oxford University Press.

Quraishi, I. H., Benjamin, C. F., Spencer, D. D., Blumenfeld, H. & Alkawadri, R. (2017). Impairment of consciousness induced by bilateral electrical stimulation of the frontal convexity. *Epilepsy and Behavior Case Reports*, *8*, 117–122.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*(1–3), 17–42.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K. & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*(7451), 585–590.

Rolls, E. T. (2004). A higher order syntactic thought (HOST) theory of consciousness. *Advances in Consciousness Research*, *56*, 137–172.

Rosenthal, D. (1997). A theory of consciousness. In N. Block, O. Flanagan & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates*. Cambridge, MA: MIT Press/Bradford Books.

Rosenthal, D. (2002). Explaining consciousness. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 406–421). Oxford: Oxford University Press.

Rosenthal, D. (2011). Exaggerated reports: Reply to Block. *Analysis*, *71*(3), 431–437.

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E. & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, *1*(3), 165–175.

Sahraie, A., Weiskrantz, L., Barbur, J., Simmons, A., Williams, S. & Brammer, M. (1997). Pattern of neuronal activity associated with conscious and unconscious processing of visual signals. *Proceedings of the National Academy of Sciences*, *94*(17), 9406–9411.

Schwitzgebel, E. (2012). Introspection, what? In D. Smithies & D. Stoljar (Eds.), *Introspection and consciousness*. Oxford: Oxford University Press.

Sebastián, M. Á. (2013). Not a HOT dream. In R. Brown (Ed.), *Consciousness inside and out: Phenomenology, neuroscience, and the nature of experience*. New Yor, NY: Springer.

Shekhar, M. & Rahnev, D. (2018). Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. *Journal of Neuroscience*, *38*(22), 5078–5087.

Siegel, S. (2006). Subject and object in the contents of visual experience. *The Philosophical Review*, *115*(3), 355–388.

Siewert, C. (1998). *The significance of consciousness*. New Jersey: Princeton University Press.

Simons, D. J. & Ambinder, M. S. (2005). Change blindness: Theory and consequences. *Current Directions in Psychological Science*, *14*(1), 44–48.

Simons, D. J. & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.

Symonds, C. & MacKenzie, I. (1957). Bilateral loss of vision from cerebral infarction. *Brain*, *80*(4), 415–455.

Szczepanski, S. M. & Knight, R. T. (2014). Insights into human behavior from lesions to the prefrontal cortex. *Neuron*, *83*(5), 1002–1018.

Turatto, M., Sandrini, M. & Miniussi, C. (2004). The role of the right dorsolateral prefrontal cortex in visual change awareness. *Neuroreport*, *15*(16), 2549–2552.

Tye, M. (2002). *Consciousness, color, and content*. Cambridge, MA: MIT Press.

Vaccaro, A. G. & Fleming, S. M. (2018). Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain and Neuroscience Advances*, *2*, 2398212818810591.

Van Gulick, R. (2004). Higher-order global states (HOGS): An alternative higher-order model. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness*. Amsterdam and Philadelphia: John Benjamins.

Voytek, B., Davis, M., Yago, E., Barceló, F., Vogel, E. K. & Knight, R. T. (2010). Dynamic neuroplasticity after human prefrontal cortex damage, *68*(3), 401–408.

Walsh, V., Ellison, A., Battelli, L. & Cowey, A. (1998). Task–specific impairments and enhancements induced by magnetic stimulation of human visual area V5. *Proceedings of the Royal Society of London B: Biological Sciences*, *265*(1395), 537–543.

Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366–383.

Weisberg, J. (1999). Active, thin and hot! An actualist response to Carruthers' dispositionalist HOT view. *Psyche*, *5*(6), 1–7.

Wolfe, J. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge, MA: MIT Press.

Xu, Y. (2017). Reevaluating the sensory account of visual working memory storage. *Trends in Cognitive Sciences*, *21*(10), 794–815.

Zeki, S. (1990). A century of cerebral achromatopsia. *Brain*, *113*(6), 1721–1777.

Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Sciences*, *7*(5), 214–218.

Zihl, J., Von Cramon, D. & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain*, *106*(2), 313–340.

How to cite this article: Kozuch B. Underwhelming force: Evaluating the neuropsychological evidence for higher-order theories of consciousness. *Mind & Language*. 2021;1–24. https://doi.org/10.1111/mila.12363