

## *Empty Higher Order States in Higher Order Theories of Consciousness*

SINEM ELKATIP HATIPOGLU  
*Marmara University, Istanbul, Turkey*

*According to higher order (HO) theories of consciousness, a mental state is conscious when there is a HO state about it. However, some HO states do not seem to be about other existing mental states. It is possible to resolve this problem since targetless HO states resemble HO states that misrepresent but the assumption that HO states always target other existing mental states is at odds with the theory since HO states are not only necessary but also sufficient for phenomenal consciousness according to the theory. Given the sufficiency of the HO states for consciousness, there is a need to understand the emergence of HO states as a non-random phenomenon to avoid the difficulties caused by targetless HO states. I suggest it is possible to develop such an understanding by thinking of HO states as predictive states in accordance with the predictive processing theory of the mind.*

**Keywords:** Consciousness; higher order theories; empty higher order states; predictive processing.

### *Introduction*

According to higher order (HO) theories of consciousness, a mental state is conscious when its subject is aware of it in a suitable manner. Among different accounts of this awareness (see Gennaro 2004 for an overview), one view states that the awareness involves a mental state that is distinct from the mental state one gets to be aware of. For instance, according to Rosenthal's (2005) higher order thought (HOT) theory, a mental state is conscious when its subject is aware of the state by way of having thought about it. The mental state one gets to be aware of is the target state or the lower order (LO) state and the thought about it is the HO state.

It is important to note that according to the HOT theory, HO states are not only necessary but also sufficient for phenomenal consciousness. I will refer to this as the *sufficiency principle*. Accordingly, the subject does not necessarily have to be in some LO state for the HO state to represent its subject to be in that LO state. Also, even if the subject is in some mental state, it does not necessarily follow that she will be phenomenally conscious of it in the absence of a HO state. This “division of phenomenal labor” between the LO and the HO state has been a source of criticism directed at HO theories.<sup>1</sup>

Higher order states may accurately represent the mental state that the subject is in, or misrepresent it, or represent the subject to be in some mental state that she is not even in. Criticism of the division of phenomenal labor is particularly powerful in this last case, viz., the case of empty HO states where there is a HO state without a target state (see for instance the discussion between Block 2011a, 2011b, Rosenthal 2011 and Weisberg 2011a and 2011b). I refer to this criticism as *the empty HO state objection*. In the case of empty HO states, there is an additional concern about which particular state is conscious in virtue of the HOT since the possibility of an empty HO state shows that the theory is committed to saying that subjects can be phenomenally conscious of mental states that they are not in.

Wilberg (2010) emphasizes this particular problem when he raises the question of which existing token mental state is conscious in virtue of the empty HOT and finds Rosenthal’s suggestion that the conscious mental state “... may be a merely notional state and may not actually exist” (2000: 232) to be in conflict with the fact that his theory is a theory of *state* consciousness, according to which consciousness would be a property of a mental state token. Wilberg denies that consciousness is only a matter of appearance and consequently denies that when it seems a certain way to a subject then she must be in a conscious state. Otherwise, one would be forced to simultaneously say that the mental state token exists and does not exist in the case of empty HO states. The mental state token does not exist because the HOT is empty and it exists since it seems a certain way to the subject, i.e. it seems to the subject as if she is in a specific mental state. To remove this incoherence, Wilberg’s account of empty HO states consists in what he calls the “no consciousness account” according to which a subject is not in a conscious state in the case of empty HO states.

Berger (2014) undermines Wilberg’s (2010) argument for incoherence and reinstates the notion of consciousness as a matter of appearance, more specifically as a matter of which mental state it seems to oneself to be in. Hence according to Berger, one’s awareness of a mental state strictly speaking should be understood as one’s awareness of

<sup>1</sup> This critique of higher order theories was first taken up by Byrne (1997) and then by Naender (1998), and Levine (2001). Later, others such as Kriegel (2003) and Mandik (2009) have addressed the same issue. The phrase “division of phenomenal labor” appears in Naender (1998).

oneself as being in a mental state. As such, he argues that despite the terminology of state consciousness, the property of consciousness really attaches itself to individuals (2014: 831). Therefore, Berger says that there is no problem with empty HO states if consciousness is taken as a property of subjects and not existing mental states.

Block (2011a) makes a distinction between the ambitious and the modest version of HO theories and contends that when faced with the question of why putting together an unconscious pain with an unconscious thought about it results in a conscious pain, the ambitious theory must provide a meaningful answer since unlike the modest view it aims at an account of the nature of what it is likeness. He then argues that the HOT theory cannot achieve this because it abuses the notion of what it is likeness as can be seen in its response to the empty HO state problem. Block (2011a: 426) says that “If what it is likeness is supposed to *matter* in the same way *whether it exists or not*, that just shows that ‘what it is like’ is being used in a misleading way” (his italics).

Farrell (2017) argues that if empty higher order states are endorsed by HO theories, then one should deny that these theories account for what-it-is-likeness, and without such an account a theory of consciousness is no longer an ambitious theory. According to Farrell, to undermine the problem of empty HO states and of misrepresentation for that matter, HO theorists adopt what he calls an occurrent reading of there being something it is like for the subject to be in a mental state (2017: 2748) and a loose reading of there being an occurrence of what-it-is-likeness associated with a mental state (2017: 2750). According to these readings, there being something it is like for the subject to be in a mental state entails that there is an occurrence of what it is likeness associated with that mental state but the subject does not have to be in some mental state for there to be a what it is likeness associated with that mental state. Farrell then argues that neither of these readings fit with our ordinary conception of consciousness based on the Nagelian definition and therefore HO theorists either would not really be responding to their opponents’ arguments in adopting these readings or they become non-ambitious theories of consciousness since they cannot provide an account of what-it-is-likeness.

Gennaro (2012) tries to resolve the issue by developing another version of the HOT theory, viz., WIV (Wide Intrinsicity View) theory, according to which the HO state is actually a part of the lower order state and together they form a complex conscious state, hence the LO state is not numerically distinct from the HO state.

I contend that the empty HO state objection arises as a consequence of not taking the sufficiency principle seriously enough and relies on the false assumption that a HO state must target a LO state. As Rosenthal (2000: 232) points out, the so-called LO state can be a non-existent or a notional state. However, the sufficiency principle is not welcomed because there is not enough literature discussing the emergence of HO

states or how they may be related to LO states when and if they are related to them.<sup>2</sup> Thus, the emergence of the HO states seems like a random phenomenon that further fuels the empty HO state objection. By providing some theory about the emergence of HO states, both the dichotomy between the LO and the HO state and the sufficiency principle would be better understood. While the HO theory that I focus on is Rosenthal's HOT theory, most of the things discussed here are relevant to any HO theory where the HO state is distinct from the LO one.

In this paper, I first discuss a way to undermine the empty HO state objection which relies on the arbitrariness between the empty HO state phenomenon and misrepresentation and then explain what is wrong with this approach. The right approach should be compatible with the tenets of the theory viz., the sufficiency principle. This, I suggest, is possible by taking HO states to be similar to predictive states in accordance with predictive processing theory of the mind (see Clark 2016, Metzinger and Wiese 2017). My purpose is not to develop a complete theory of the HO states as predictive states but only to pave the way for a theory of the emergence of HO states.

### 1. *The so-called accurately represented targets, misrepresented targets and absent targets*

Consider the following examples according to which I am in a,

- (1) LO mental state of seeing a green apple
- (2) LO mental state of seeing a green ball
- (3) LO mental state of seeing a red bowl

According to the HOT theory, it is possible for the subject to have a HOT with the content "I'm seeing a green apple" in all these cases and be phenomenally conscious of seeing a green apple. I follow Weisberg (2011a: 416) in calling a case like (1) veridical representation, (2) misrepresentation and (3) as involving a HO state with no target or an empty HO state.

Wilberg (2010) says that it is possible to understand cases of misrepresentation as cases where the target of the HOT does not exist. Similarly, Rosenthal (2004: 32) finds the "distinction between an absent target and a misrepresented target ... arbitrary" and says,

Suppose my higher-order awareness is of a state with property P, but the target isn't P, but rather Q. We could say that the higher-order awareness misrepresents the target, but we could equally well say that it's an awareness of a state that doesn't occur. The more dramatic the misrepresentation, the greater the temptation to say the target is absent; but it's plainly open in any such case to say either.

<sup>2</sup> When it comes to the relation between HO states and LO states, Rosenthal (1993a) denies it to be causal and the best scenario is that of an accompaniment. This being the case, one wonders if there is any limit—and on what grounds to the way a HO may represent a LO state.

One may then say that if misrepresentations are unobjectionable, so should empty higher order states be. However, I contend that this is not the right approach to defend the HO theory from empty HO state objection. While it may be tempting to resolve the issue about which token state gets to be conscious in the case of empty HO states by likening absent targets to misrepresented ones and thereby assigning targets to them, such an approach only reinstates the assumption that a HO state always targets a LO state. This would overlook the sufficiency principle.<sup>3</sup>

The mental state the HO state represents its subject to be in may coincide with a certain existing LO state that the subject is in but this is neither a necessary aspect of the theory nor is it a necessary feature of the relation between the HO state and the LO state, assuming there is a relation. It is interesting to note that the so called veridical cases where the notion of ‘HO state targeting a LO state’ is perhaps the most powerful may also be redescribed in a way to involve misrepresentations. For instance, in the case of (1) the subject may be phenomenally conscious of seeing the apple’s color as a generic green rather than the particular shade of green the LO state represents the apple to have. The orthodox way of thinking about this is usually as a case of veridical representation where certain subtleties are lost in the HO representation of the LO state. However, given the sufficiency principle, it is just as reasonable to think of the HO state independently of the LO state. Similar to Rosenthal’s earlier suggestion one might say that the subject’s HO awareness is of a state with property P (generic green) but that the target isn’t P but rather Q (the particular shade of green). Hence one might suggest that there is a certain sense of arbitrariness concerning the distinction between veridically represented LO states, misrepresented LO states and absent LO states. While that may be true, it would be wrong to use this idea and contend that there are no empty HO states to undermine the empty HO state objection since the idea relies on the false assumption that a HO state must always target an existing LO state.

## 2. *Randomness and empty higher order states*

Resembling the empty HO state phenomenon to misrepresentation and thereby rendering the HO state non-empty reinstates the idea that a HO state must always target a LO state and therefore is at odds with the sufficiency principle. If the sufficiency principle is dispensable for HO theories, then the above approach might work but I don’t think it is dispensable. Hence the HO theorist needs to address why the suf-

<sup>3</sup> I’m not suggesting that Rosenthal’s purpose (2004: 32) in the quotation above is to undermine the empty HO state objection based on the arbitrariness. Instead it should be understood as providing some clarification on the notion of an absent target or an awareness of a state that doesn’t occur.

iciency principle which is really at the core of the empty HO state objection is not welcomed.

As Gennaro (2012: 60) and before him Levine (2001: 108) have discussed,<sup>4</sup> since HO states are sufficient for phenomenal consciousness, there seems to be no point of there being a lower order state, especially a numerically distinct one.<sup>5</sup> The presence of an actually existing LO state is possible but neither necessary nor sufficient for consciousness and empty HO states stand out because they make the sufficiency of HO states for consciousness more obvious.

Without articulating why and how HO states come about, an inevitable sense of randomness threatens the theory. The concept of an absent target makes this randomness obvious, while a misrepresented target promises a story about how the HO state is still about the LO state and fits better with our general understanding of mental lives by making them seem less random. It seems that this sense of randomness fuels the empty HO state objection. Therefore, one could be tempted to argue that there are no genuinely empty HO states but only misrepresented targets. However, as mentioned before, I consider this at odds with the very tenets of the theory.<sup>6</sup>

If randomness is to be avoided and some theoretical background is to be provided for the emergence of HO states, I'd like to suggest that this is possible by incorporating the mental history of the subject into the emergence of HO states. While LO states may be a part of that history, they would not stand out in any special way in terms of their relation to the HO states.

I will not try to articulate in detail what the mental history of a subject refers to but it is meant to be the kind of thing that gives rise to the phenomenal differences between for instance Mary's<sup>7</sup> first experience of seeing a red chair after leaving the black and white room and her

<sup>4</sup> Gennaro (2012: 60) says that one faces the question of "... what the point of having both a LO and HO state is if only one of them determines the conscious experience." Likewise, Levine (2001: 108) says that "the first-order state plays no genuine role in determining the qualitative character of experience."

<sup>5</sup> This is probably why Gennaro (2012) develops his version of a HO theory of consciousness, viz. WIV (Wide Intrinsicity View) theory according to which the higher order state is a part of the lower order state and together they form a complex conscious state, hence the LO state is not numerically distinct from the HO state. My purpose is to assess the empty HO state objection for theories where the LO state is numerically distinct from the HO state, not when it is a part of the HO state. Obviously, the question whether WIV is able to tackle the empty HO state objection while remaining to be a higher order theory is worth examining but I cannot undertake this task here.

<sup>6</sup> Besides, even if absent targets are replaced by misrepresented ones, one still faces the question of why HO states would misrepresent their targets in this radical way or in what sense a HOT with the content 'I'm seeing a green apple' would still be about 'perception of red bowl' LO state.

<sup>7</sup> Jackson's (1986) example of the super scientist who is omniscient concerning physical knowledge and knows all about colors but was grown up in a black and white room and has never seen a colored object before.

experience of the same red chair two years after she leaves the room. These differences, while taken for granted, are not addressed sufficiently with the exception of Rosenthal (1991: 33-4, 2002: 413-4) who argues that one of the advantages of the HOT theory is in its ability to explain how one's conceptual resources influence the phenomenological features of one's experiences since the HO state is a thought.<sup>8</sup>

Consider the following example of the impact one's mental history has on one's consciousness. A woman is sitting in the lobby of a building, waiting anxiously to meet her long-lost brother. She is constantly checking the sliding doors that open to the lobby. Then a strong wind causes a plastic bag to fly in. It is conceivable that being phenomenally conscious of seeing her brother, the woman gets off her seat to meet him and soon realizes it was just a plastic bag. Perhaps one might suggest that given the 'perception of the plastic bag flying in' as the LO state, along with the desire to see the brother, the anticipation etc., the brain is in some sense forced to predict that her brother has arrived resulting in the HOT 'I'm in a mental state of seeing my brother.' One might even suggest that this prediction is for the organism's well-being, for instance, to momentarily reduce the stress the subject suffers from. Hence even though the HO state is targetless, there is a certain background, a certain mental space in which this particular HO state comes about.

This is a direction in which HO theories may further be developed, viz., by providing an account according to which empty HO states arise for the organism's well-being given its mental history. This would eliminate randomness for two reasons. Firstly, empty HO states would be driven by a purpose. This purpose could be to sustain a certain level of equilibrium in the subject's mental life by avoiding too much stress. It could be a reaction to the mental history of the subject. Just as blinking is a physical reaction to protect one's eye when something gets close to it, empty HO states could be a mental reaction to protect one's mental health under conditions where the subject needs to have an experience  $x$  even though she is not in that particular mental state  $x$ . Secondly, empty HO states would be grounded in some mental space rather than being randomly generated since they arise in relation to the subject's history.

In fact, it is possible to think this way about HO states in general and not just empty ones. One way to do this is to think of HO states as the predictive states in accordance with the predictive processing theory of the mind (PPT) (see Clark 2016, Hohwy 2013, Metzinger and Wise 2017). Given the subject's history, the HO state's representation of its subject to be in some mental state would actually be a prediction of what the subject would be phenomenally conscious of. PPT emphasizes the constructive nature of mental episodes, such as perception and the top-down processing that is involved. Hence perception is not merely

<sup>8</sup> He gives the example of wine tasting, musical experience (1991: 33-4) and the experience of hearing the sound of an oboe (2002: 413-4).

passive and stimulus-driven. Instead, it is active and also hierarchical. This top-down processing is not something that is effective only when sensory input cannot be relied on but it is essential to and constructive of perception. Put simply, the brain makes use of computational models in accordance with Bayesian inference as a computational method to make predictions about the external world that the subject is in and the possible causes of the effects that the subject is receiving information about through sensory signals. A more dramatic way to put this is to say that the brain dreams in a world where dreaming is not random but very much controlled (Metzinger 2003: 52).

The next step, again put simply, involves the brain asking to itself if the prediction it's made is correct. This is done by taking into account the sensory input and checking if there is a mismatch between the sensory input and the prediction, provided that the sensory input is reliable. If the sensory input is not to be trusted, i.e., if it is too noisy or ambiguous, even if there is a mismatch, the sensory input is undermined and prevented from being further processed. However, if the sensory input is reliable and there is a mismatch between it and the prediction, the computational model that the brain uses to make its predictions is revised to decrease errors in future predictions.

Given the sufficiency principle and the relevance of conceptual resources to one's phenomenal consciousness in HOT theory, granted that the HO state is a thought, I contend that the HOT theory of consciousness is the most compatible one with PPT since it allows for the top-down process that PPT endorses rather than a bottom-up process. Interestingly enough, the evidence for this lies in the phenomenon of empty HO states even though empty HO states are usually the source of an objection to HO theories, as discussed in the beginning. The simple fact that being in a conscious state does not necessarily involve being in that state in the HOT theory may be seen as evidence for the top-down process. Just as the sensory input in perception according to predictive processing is used to check if the prediction is correct, and therefore not initially essential to the prediction in the top-down framework, the mental state that the subject is in may be considered to be non-essential to the HOT about it but may later be used to check the accuracy of the HOT, that is if the subject is indeed in such a mental state.

So instead of the subject being in a mental state and there being a HO representation of that mental state the subject is allegedly in, which would be a bottom-up process, the HO representation can be taken to be a prediction of the mental state the subject would be in regardless of whether or not the subject is in that state. Hence technically, the HO representation would not be the representation of a LO state strictly speaking but a thought of a predictive nature about some LO mental state that the subject might be in, given the circumstances. And again in accordance with PPT, the next step would involve checking if the prediction is correct. In the example given above, since the subject does not see her brother in the moments that follow as she approaches the



doors, the prediction would need to be corrected. If empty HO states are typically taken to be rare or not to last long, this prediction-checking followed by a revision when needed would provide an explanation for the rarity or short duration of empty HO states. Another analogy that may be observed between PPT and HOT theories is that just as the predictions are not experienced as predictions by the subject, the HOTs are not typically conceived as thoughts that the subjects are conscious of having.<sup>9</sup>

This way of thinking about the emergence of empty HO states, or HO states in general calls for a change in our ordinary ways of thinking about consciousness which usually include a bottom up process of being in a mental state and then being aware of being in it. However, as Rosenthal (2004: 41) notes, consciousness is not actually about being in a state and being conscious of being in it. The first part of this conjunction is in fact somewhat irrelevant to the second part. Studies in predictive processing have paved the way for this top-down framework and there is no obvious reason to refute a similar framework in theories of consciousness. As Metzinger (2003: 52) also says,

[A] fruitful way of looking at the human brain, therefore, is as a system which, even in ordinary waking states, constantly hallucinates at the world, as a system that constantly lets its internal autonomous simulational dynamics collide with the ongoing flow of sensory input, vigorously dreaming at the world and thereby generating the content of phenomenal experience.

Undoubtedly further work on how HOT theory of phenomenal consciousness and PPT can be brought together is needed and for reasons discussed this seems to be a promising way to enhance our understanding of the mind and of consciousness.

## References

- Berger, J. 2014. "Consciousness is not a property of states: A reply to Wilberg." *Philosophical Psychology* 27 (6): 829–842.
- Block, N. 2011a. "The higher-order approach to consciousness is defunct." *Analysis* 71 (3): 419–431.
- Block, N. 2011b. "Response to Rosenthal and Weisberg." *Analysis* 71 (3): 443–448.
- Byrne, A. 1997. "Some like it HOT: Consciousness and higher order thoughts." *Philosophical Studies* 86 (2): 103–129.
- Clark, A. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.
- Farrell, J. 2017. "Higher-order theories of consciousness and what-it-is-like-ness." *Philosophical Studies* 175 (11): 2743–2761.
- Gennaro, R. J. (ed.). 2004. *Higher-Order Theories of Consciousness: An anthology*. Amsterdam: John Benjamins Publishing.
- Gennaro, R.J. 2012. *Consciousness Paradox*. Cambridge: MIT Press.
- Gennaro, R. J. 2017. *Consciousness*. New York: Routledge.

<sup>9</sup> Unless the HOT itself is conscious which is rarely the case, if possible.

- Hohwy, J. 2013. *The Predictive Mind*. Oxford: Oxford University Press.
- Jackson, F. 1986. "What Mary didn't know." *Journal of Philosophy* 83 (5): 291–295.
- Kriegel, U. 2003. "Consciousness as intransitive self-consciousness: Two views and an argument." *Canadian Journal of Philosophy* 33: 103–132.
- Kriegel, U. 2004. "Consciousness and self-consciousness." *The Monist* 87 (2): 182–205.
- Levine, J. 2001. *Purple Haze: The Puzzle of Conscious Experience*. Cambridge: MIT Press.
- Neander, K. 1998. "The division of phenomenal labor: A problem for representational theories of consciousness." *Philosophical Perspectives* 12: 411–434.
- Mandik, P. 2009. "Beware of the unicorn: Consciousness as being represented and other things that don't exist." *Journal of Consciousness Studies* 16 (1): 5–36.
- Metzinger, T. 2003. *Being No One: The Self-Model Theory of Subjectivity*. Cambridge: MIT Press.
- Metzinger, T. and Wiese W. (eds.) 2017. *Philosophy and Predictive Processing*. MIND Group.
- Rosenthal, D. 1991. "The Independence of Consciousness and Sensory Quality." *Philosophical Issues* 1: 15–36.
- Rosenthal, D. 1993a. "Higher order thoughts and the appendage theory of consciousness." *Philosophical Psychology* 6 (2): 155–166.
- Rosenthal, D. 1993b. "State consciousness and transitive consciousness." *Consciousness and cognition* 2 (4): 355–363.
- Rosenthal, D. 2000. "Metacognition and higher order thoughts." *Consciousness and cognition* 9: 231–242.
- Rosenthal, D. 2002. "Explaining consciousness." In D. Chalmers (ed.). *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press, 406–421.
- Rosenthal, D. 2004. "Varieties of higher order theory." In R. Gennaro (ed.). *Higher-Order Theories of Consciousness: An anthology*. Amsterdam: John Benjamins Publishing, 17–44.
- Rosenthal, D. 2005. *Consciousness and Mind*. Oxford: Oxford University Press.
- Rosenthal, D. 2011. "Exaggerated reports: A reply to Block." *Analysis* 71 (3): 431–437.
- Weisberg, J. 2011a. "Misrepresenting consciousness." *Philosophical Studies* 154: 409–433.
- Weisberg, J. 2011b. "Abusing the notion of what-it-is-like-ness: A response to Block." *Analysis* 71 (3): 443–448.
- Wilberg, J. 2010. "Consciousness and false HOTs." *Philosophical Psychology* 23 (5): 617–638.