

Essentially Indexical Higher-Order Thoughts and Dispositions of Inexistent Mental States

Asger Kirkeby-Hinrup

Abstract

In a recent paper, Miguel Sebastián has posed an objection to the higher-order thought theory of consciousness. The higher-order thought theory of consciousness holds that a mental state is conscious when it is the intentional object of a higher-order thought (HOT). Often, the HOT is conceived to be essentially indexical. This means that the HOT somehow picks out the individual who tokens it. Sebastián argues that the way David Rosenthal conceives of the essential indexicality of higher-order thoughts leads to a vicious regress. The regress is vicious, or so it is argued, because our cognitive capacities cannot cater to infinite strings of HOTs. Therefore, according to Sebastián, the higher-order thought theory fails to provide a satisfactory account of consciousness. In this paper, I argue that the way Sebastián sets up the regress argument relies on the implausible view that inexistent mental states have real dispositions on par with the dispositions of ‘real’ occurrent mental states. This means that the regress cannot get off the ground. In addition to this, I argue that the fact that there are limits to the cognitive capacities of humans is not in itself an argument against the HOT theory. Despite dismissing Sebastián’s regress argument, I see value in the discussion of the indexicality criterion, and to precipitate this I consider the possibility that the indexicality of HOTs may be transitive, and explore possible ways to ground such a transitive conception of indexicality.

1. Introduction

In a recent paper, Miguel Sebastián (2018) has leveraged a new objection to the higher-order thought theory of consciousness (HOTTC). The HOTTC posits that the way individuals become aware of themselves as being in (viz. are conscious of) a mental state *C* is by having a higher-order thought (HOT), whose intentional object is *C*. Different versions of HOTTC involve various claims about the proper nature of the HOT, if it is to render the individual conscious of *C* (see e.g. Brown, 2015; Coleman, 2015; Gennaro, 2006;

Matey, 2006; Weisberg, 1999). Sebastián focuses on one widely shared claim, namely the idea that the HOT must be essentially indexical. Sebastián argues that this leads to a vicious regress of infinite dispositions. The regress argument specifically targets the variant of HOTTC developed by David Rosenthal (e.g. Rosenthal, 1997, 2002a; Rosenthal, 2011). Sebastián's objection turns on Rosenthal's explication of essential indexicality in terms of a disposition to identify the individual tokening a HOT (Rosenthal, 2004, 2011). Sebastián concludes from his objection that the vicious regress shows that Rosenthal's HOTTC fails to provide a satisfactory account of consciousness (Sebastián, 2018, p. 9).

I will argue that Sebastián's argument against the HOTTC does not succeed. The reason is that the vicious regress depends on the implausible view that inexistent mental states have existing dispositions. Despite my criticism of the regress argument, I consider discussion of the indexicality criterion relevant and valuable to the debate on higher-order theories. Not the least because the indexicality criterion has received relatively little attention compared to other issues. Thus, I see Sebastián's objection as valuable simply for shedding light on this aspect of the HOTTC. For this reason, I devote the second part of the paper to exploring an alternative way of conceiving of the indexicality of HOTs. This way consists in suggesting that what is picked out by the essentially indexical property of a HOT, i.e. the individual tokening the HOT, is transitive.

In the next section, I will provide a brief overview of the HOTTC and summarize Sebastián's regress argument. In section three, I present what I view as the central problem with the regress argument. In the fourth section, I explore the possibility that the indexicality of HOTs may be transitive and consider four different ways one may ground transitive indexicality. Finally, in section five, I offer some concluding remarks.

2. HOTTC, Essential indexicality, and the proposed vicious regress

As noted above, HOTTCs come in many variants depending on how one conceives of the relationship between a HOT and its target, as well as the properties of the HOT. However, given that Sebastián specifically targets the HOTTC proposed by David Rosenthal, I will limit myself to summarizing this version here.

HOTTC theories of consciousness proceed from the intuitive idea that a mental state, one is in no way aware of being in, is not a conscious state. This is usually taken to form the foundation for the so-called *Transitivity Principle*

(TP), which is the idea that a conscious state is a state one is aware of oneself as being in (see e.g. Matey, 2006; Rosenthal, 1997; Weisberg, 2010).

The way one becomes aware of being in a given mental state **m**, on the HOT account, is by having a suitable thought about **m**. For instance, Rosenthal (2004, p. 160) says "...a mental state's being conscious consists in its being accompanied by a suitable thought that one is, oneself, in that state". Traditionally, several riders have been attached to the properties of the HOT and how it comes about. These riders serve to pinpoint the kind of properties a HOT should have, in order for it to render an individual aware of herself as being in **m**. Such riders, for instance that – in the normal case – the HOT is itself an unconscious state, are central parts of the higher-order doctrine. That the HOT is normally itself an unconscious state allows the HOTTC to explain consciousness by reference to something unconscious, thereby avoiding circularity. Additionally, if a HOT was necessarily conscious, this would yield a regress. To be conscious, a HOT would require another HOT, which would need another, and so forth.

I shall not rehearse all the riders here (see e.g. Rosenthal, 1997, 2002a, 2002b, 2004 for details.). For the present purposes, the relevant rider on which to focus is the indexicality criterion. The indexicality criterion serves to identify the tokener of a HOT as the individual who is in the mental state the HOT is about. As Rosenthal puts it: "A HOT makes one conscious of oneself as being in a particular mental state because it has the content *that one is, oneself, in that state*" (Rosenthal, 2004, p. 165, emphasis added). The way the HOT achieves this is by being *essentially indexical*. Essential indexicality – in this context – means that a critical feature of consciousness rendering HOTs is that they refer to the individual who tokens them. That a HOT is essentially indexical does not mean that only individuals with indexical *concepts* can have HOTs. Rather, this means that the HOT implicitly identifies the individual who is in the state that the HOT represents. This identification intuitively seems highly relevant to a notion of consciousness conceived according to the transitivity principle, since this handles the identification of *oneself*. (see e.g. Rosenthal, 2004, 2011 for additional details of the indexicality criterion).

Sebastián takes his point of departure in Rosenthal's (2011) explication of how to conceive of the essential indexicality of HOTs. Rosenthal holds that thinking of the essential indexicality in terms of an occurrent property of a HOT will not do. Instead he suggests:

“Though the HOT does not describe that individual as the thinker of the HOT, the individual is disposed to do so should the question ever arise. [...] the question seldom if ever does arise; so the individual that has the HOT may never actually perform that identification. But the disposition to do so constitutes a tacit identification of the self that the HOT ascribes pain to. And that constitutes the essentially indexical self-reference.”

(Rosenthal, 2011, p. 30)

Following Lewis, Sebastián proposes to analyse Rosenthal’s deployment of dispositions in terms of stimulus and manifestation conditions (e.g. Sebastián, 2018, p. 5). Stimulus conditions are conceived of as the conditions that are necessary and sufficient for a disposition to be realized e.g. become manifest. Manifest conditions, in turn, are whatever happens when the stimulus conditions are met. With respect to the essential indexicality of a HOT, the stimulus condition is the subject forming a HOT with approximately the content ‘it is I, who thinks/thought that X’. The manifestation condition is the individual having a conscious state with (approximately) the content ‘it is I, who thinks/thought that X’.

From this Sebastián argues that a regress ensues. He says:

“[...] if one is to identify the individual the HOT refers to as the individual who has the HOT, then it has to be possible that the HOT becomes conscious. According to the theory, this would require it to be possible that the subject has an unconscious third-order thought to the effect that one oneself is having the HOT [...] The third-order thought deploys the first-person concept, which refers to oneself in an essentially indexical way. Therefore, the subject has the disposition to identify the individual the third-order thought refers to as the individual who has that third-order thought. But if the subject is to have such a disposition, then, by the reasoning above, it has to be possible for the third-order thought to be conscious. This in turn requires a fourth-order thought [...] and so on: ad infinitum.”

(Sebastián, 2018, p. 5)

In addition to showing there is a regress, Sebastián aims to show that the regress is vicious. In order to show this, he proposes that “the capacity to undergo phenomenally conscious experiences, requires an arbitrarily tall

hierarchy of dispositions to have HOTs.” (Sebastián, 2018, pp. 5–6). This, he concludes, makes the HOT theory depend on the question of whether creatures such as us, have the cognitive capacities to entertain an infinite string of HOTs. Reasonably, being finite creatures with limited cognitive capacities, humans cannot entertain an infinite string of HOTs. Thus, he concludes (Sebastián, 2018, p. 9) that the HOT theory “fails to provide a satisfactory account of consciousness”.

3. Inexistent states cannot have ‘real’ dispositions

In this section, I highlight a problem for Sebastián’s regress argument. This problem, I submit, is sufficient to dismantle the objection to HOT theories. Sebastián frames his objection in two steps. The first step is an attempt to establish an infinite regress of higher-order dispositions. The second is to argue that the cognitive capacities of humans are insufficient to entertain this infinity of higher-order states¹. However, in the way Sebastián sets up his regress argument, he treats the disposition of an actual occurrent state and the disposition of an inexistent HOT about that state equally. Treating inexistent dispositions on par with actual or ‘real’ dispositions poses a problem for both steps in Sebastián’s argument.

Let us start with an example, to illustrate the kind of dispositions that are at stake here. At any given time, there will be just one occurrent conscious thought. It is, *ex hypothesis*, true that the occurrent state will have an *actual* disposition. That actual disposition will have stimulus and manifestation conditions in accordance with Sebastián’s analysis. For instance, suppose I am in a conscious state **C1** with the approximate content ‘Wow, a butterfly!’². Now, I am conscious of **C1** in virtue of the presence of a HOT (**C2**) about my seeing a butterfly. This HOT has an essentially indexical property that Rosenthal cashes out as a disposition to identify the individual (me) that is excited about seeing a butterfly. The stimulus condition for **C2**’s disposition is me forming yet another HOT (**C3**) with the approximate content ‘It is I, who was excited to see a butterfly’ or ‘the one who was

¹ Also, it is worth noting that we can devise plenty of hypothetical conscious thoughts with dispositions for further thoughts, for which we can define clear stimulus and manifestation criteria that we cannot carry out in practice. For instance, children trying to count to the ‘highest number’.

² I am aware that there are more stringent ways of cashing out the propositional content of conscious states e.g., “I am having a visual representation of a butterfly”. For the sake of exposition, grant me this looser way of speaking.

conscious of seeing that butterfly was me'. The manifestation conditions for the disposition of **C2** is me actually forming **C3** resulting in my having a conscious thought with the content described. Now, and this is the crucial part, in the scenario just described **C3** does not exist. **C3** is merely the manifestation condition of a disposition. It is, *ex hypothesis*, counterfactually true that *if C3 existed, C3 would have a disposition (with the manifestation condition of me forming C4) that would allow me to identify the individual (again, me) that was conscious of C3.*

How should we think of this counterfactual disposition of **C3**? If we treat it as a real disposition with ontological commitments identical to those we confer on **C2**, we are committed to treating **C4**... **Cn** in the same way, and the regress follows. This problem is similar to another issue Rosenthal has addressed previously. Arguing against the idea that all mental states are conscious, Rosenthal (*passim*) acknowledges that if every HOT, *qua* mental state, is conscious and he wishes to explain a state being conscious in terms of the *presence* of a HOT about it, this leads to an infinite regress. But importantly, the regress proposed by Sebastián differs from this example in so far as the disposition of **C3** does not exist, i.e. it is not *present*, whereas the HOTs in the regress Rosenthal discusses *are* present, i.e. exist. Basically, the reason the regress argument fails is that *inexistent* things cannot have *real or existing* properties, not even dispositional ones, because they do not exist (see e.g. Mandik, 2009; and Weisberg, 2010 for other discussions of non-existent properties in relation to the HOTTC.). This means the regress never gets off the ground.

Turning to the second step of Sebastián's argument, let us consider the question of whether the capacities of our cognitive system can provide the support it is intended to. It seems they cannot because all that matters with respect to our cognitive capacities in relation to the HOTTC is that the *existing* disposition of an *occurrent* HOT, in principle, *can* be realized. So, the question then is whether humans *in fact* ever reach a level of higher-order states that we cannot cognitively entertain. When put in this way, it is clear, that humans do not because it is stipulated in the question that they *cannot*. So, a better way of putting this is by 1) asking whether humans ever *approach* this limit, and 2) whether merely approaching it is a problem. Regarding 1), the question of whether humans approach the limit, it seems this is already covered by Rosenthal when he highlights (in the passage quoted in the previous section) that the individual is disposed to if the question should ever arise, and that the questions seldom if ever arises. Thus,

it seems that what Sebastián has proposed as a problem is merely a variant of the possibility of endless introspective states that have already been dealt with and debunked by Rosenthal in earlier work. For instance, Rosenthal says “Neural implementation is not a problem, since ample cortical resources exist to accommodate *actual* HOTs. And, though introspection seems to suggest that the mind cannot accommodate very many actual HOTs at a time, that worry is also groundless. Introspection can tell us only about our conscious states, and by hypothesis HOTs are seldom conscious” (Rosenthal, 2002a, p. 410, my emphasis). And similarly, in a much earlier paper Rosenthal says: “Indeed, we would expect, instead, that the third-order thoughts that confer consciousness on such second- order thoughts would be relatively rare; it is hard to hold in mind a thought about a thought that is in turn about a thought. So, the present account correctly predicts that we would seldom be aware of our second-order thoughts, and this actually helps confirm the account.” (Rosenthal, 1986, p. 336). Implicit in Rosenthal’s view here is that even more rarely will we be aware of our third-order thoughts, and so forth. As for 2), the question of whether approaching the level is a problem, most people will recognize that when we try to introspect our introspective states the content becomes more and more diffuse and hard to entertain. The thoughts simply stop making sense beyond a certain point. This introspective phenomenon actually may point to real limitations of our cognitive machinery, but it does not suffice to debunk the HOTTC. The fact that our cognitive machinery has limitations is not itself an argument against the HOTTC. In principle, my next thought can always be ‘it was I, who just thought X’, regardless of what X is. That is all that matters and is enough to fix indexicality. Even if this thought makes little sense to me (e.g. because of nested layers of self-reference in the successive introspections case), it does not show that HOTTC is faulty.

To summarize, Sebastián’s regress argument turned on two claims. The first claim was shown to rely on the assumption that we are allowed to treat counterfactual dispositions of inexistent entities on par with actual dispositions of existing entities. It should be clear that including inexistent entities when counting instances of an entity is either begging the question against the HOTTC or, at the very least, comes with very non-standard ontological commitments. The same goes for treating as real the counterfactual causal (dispositional) powers of inexistent entities, whether by themselves or for purposes of counting entities. Sebastián’s second claim suggested that limitations in the cognitive capacities of humans, when

coupled with the regress, shows that the HOTTC is implausible. I agree (as I think most people reasonably should) that there are in fact limitations to the cognitive capacities of humans. However, the fact that humans have limited cognitive capacities is not in itself an argument against the HOTTC. This limitation is only a problem if a theory entails that we exceed it. The regress argument was supposed to show that the HOTTC entailed this. However, the regress argument fails. To boot, according to Rosenthal, we rarely (if ever) form enough HOTs to come close to reaching the limits of our cognitive machinery.

4. Grounding indexicality

In this section, I explore an alternative way of conceiving of the indexicality criterion which suggests that the essential indexicality of HOTs is transitive. In this context, I mean by transitivity that the indexicality of a HOT is really a property of some external X to it that the HOT acquires by standing in some relation to X. Traditionally, one would say that if A is related to B and B is related to C, then A is transitively related to C. I suggest that the indexicality may be transitive (as opposed to e.g. ‘external’) exactly to leave open the possibility that in a case where multiple HOTs are instantiated, that each HOT will acquire its indexical aspect from the state that it is about, i.e. that the disposition to identify the tokener of the thought realized by e.g. C4 is acquired by C4 in virtue of being about C3, which in turn got it in virtue of being about C2³. Now, if HOTs acquire indexicality in this way, we still face a question about how the ‘bottom layer’ of the hierarchy acquires it. I will return to this shortly. For now, the crux of the idea is that the dispositions of C1, C2...Cn to identify the tokener of the previous thought are all grounded in – or refer to – something else than the HOT(s). This means that the regress can never get off the ground because this would consist in double counting (or rather infinite counting) the same disposition. One argument for holding a view on which the indexicality is transitive (or external, see footnote 3) is

³ It is of course possible that both C2, C3...Cn may refer directly, as it were, to the grounding instead of the indexicality propagating up through the representational hierarchy, in which case ‘external indexicality’ may be a more suitable description (however and importantly, this does not threaten the above reply to Sebastián’s regress). While ‘external indexicality’ is also a live option in the debate, I will allow myself to cast the discussion here in terms of transitivity. Please note that much of what I have to say about grounding in the following may be applicable even if we are dealing with external indexicality, instead of transitive indexicality. I am grateful to an anonymous reviewer for pushing for a clarification of this.

that it avoids an unnecessary proliferation of content in the cognitive system. There are obvious reasons related to cognitive economy to avoid such proliferation. Basically, if a HOT is taken to carry much of the same content as its target state, the cognitive system would have to represent that content in two locations (assuming that a HOT and its target are distinct neural states, which seems to be entailed – or at least suggested – by most HOTTCs). Since representation in human cognition, qua being a process of a dynamic biological system, requires spending depletable resources, avoiding unnecessary doubling of content is desirable. Such worries about cognitive load related to higher-order theories have been around for a while and have motivated variations of HOT theories (see e.g. Carruthers, 2000)⁴. While I am uncertain if the additional cognitive load is in fact problematic (even if HOTs do copy the content of their target states, see also discussion in the previous section), I do consider evolutionary pressure a good reason to expect cognitive economy to be a factor in how human cognition is structured.⁵ Now, Sebastián may argue that if we view indexicality as a transitive property of HOTs, the transitive property needs to be grounded in something or there will be no indexicality at all. In other words, if we pursue this solution, the indexicality of a HOT needs to piggyback on *something*. The transitive indexicality of a HOT must end up somewhere in the system, where there is ‘true’ indexicality. Sebastián (*in conversation*) seems to hold that

⁴ Sebastián, in a sense, also raises issues related to cognitive load. However, in virtue of arguing that the human cognitive system is incapable of entertaining an infinity of states, Sebastián’s version is more extreme than the one usually considered in the debates, which pertains to a finite, albeit large, number of states.

⁵ One critical issue raises its head for this sort of reply. On at least some of the so-called ‘pointer’ higher-order theories, i.e. theories where a HOT refers to, or embeds, its target rather than re-represent its content (e.g. Lau, 2019. Possibly also QHOT theory by Coleman in his 2015. But note that Coleman asserts that the content of a completed QHOT contains its embedded state, rather than merely points at it) such that the content of the HOT is “*I am in [x]*”, or “[*x*] is present” where [*x*] is a reference to the relevant target state, what the HOT adds to the picture seems to be exactly the indexicality (possibly along with assertiveness). If, on such an account, we were to view the indexicality of HOTs as transitive, which means relocating indexicality to the target state (irrespective of whether that target state is a first-order state or a HOT), what remains in terms of content for the HOT? Reasonably, if the HOT is to count in any way as a discrete mental state in its own right, it needs to have *some* content. After all, the consensus is that mental states are (at least partly) individuated by their content. This is not a critical issue for this line of reply in general but merely indicates that additional work may be required for some versions of HOT theory (but not Rosenthal’s) if they wish to pursue this line of thinking.

indexicality ultimately must ground out in a *self-concept* and unless the HOTTC can explain this self-concept it is a best an incomplete theory, or at worst *ad hoc* or inconsistent. Let us break that down into two distinct claims. The first claim is that the transitive indexicality must end up somewhere. The second claim is that a theory that deploys a transitive notion of indexicality, either must explain what grounds it, or be considered incomplete or *ad hoc*. The first claim is reasonable, and possibly even a necessary truth about (finite) transitive relations. The second claim is more tricky, and introduces a very heavy burden of proof on any theory of consciousness (under the assumption that any theory of consciousness must deal with properties, relations or states involving concepts in the vein of indexicality, self-consciousness, or self-awareness, even if it does not do so explicitly). Notice here, that there is an important difference between demanding that a theory *explains* the thing doing the grounding as opposed to merely positing *what* does the grounding, or *where* a relation grounds out, as it were. Demanding an explanation may be too heavy-handed and will throw out a lot of babies with the bathwater if adopted as a criterion for evaluating the viability of theories of consciousness⁶. Importantly, I am not attributing this claim to Sebastián. Indeed, my sense is he would agree that demanding an *explanation* is a step too far. Be that as it may, I will here be satisfied with considering the second possibility, which consists in hypothesizing about where the indexicality grounds out. However, before turning to considering the possibilities with respect to this, it is worth mentioning a separate important point of Sebastián's (*in conversation*). The point is that whatever we ground the transitive indexicality in cannot depend on (or involve) consciousness. This is an important and valid point since invoking consciousness in the grounding of indexicality would yield a circular definition, since indexicality itself was taken to be central to consciousness-rendering higher-order thoughts according to the HOTTC. With these preliminary constraints established, I will, in the rest of this section consider possible groundings for indexicality. There are quite a few possibilities, and I will not be able to give extensive detail to each of these here⁷. First, I will consider two possible

⁶ At least at the present stage of scientific understanding of the brain. Indeed, it seems likely that once we have a clearer view of the mind-brain relationship, this knowledge will naturally constrain which kinds of theories of consciousness are viable. Accounting for indexicality (or its proxies such as self-awareness), in my view, is not unlikely to figure as one such constraint.

⁷ In advance, I will mention that I do not seek to advance one possibility over the others. Nor do I purport that this is a complete list, and while I have tried to highlight advantages, constraints,

groundings for indexicality that – to my mind – can be accommodated relatively effortlessly within the existing HOTTC frameworks. After this, I consider two pieces of recent work that may be suitable to ground indexicality if either was assimilated and tweaked to fit the HOTTC framework.

Starting with two possible groundings that appear the most straightforward answers for proponents of the HOTTC. The first candidate to be the bearer of ‘true’ indexicality that grounds the transitivity of HOTs would be the lowest state in the hierarchy i.e., a first-order state. Following the reasoning above, we do not need (here at least) to provide an explanation of exactly how this indexicality is instantiated in the first-order state. However, proponents of the HOTTC would be well served to address an adjacent issue, which is that the elegance and parsimony of higher-order accounts seems threatened if the properties of first-order states and higher-order states become so different that it becomes contrived to view them as of the same ontological kind. However, a reply to this is straightforwardly available. Given that proponents of the HOTTC already embrace certain criteria a HOT must fulfil in order for it to render an individual aware of herself as being in (i.e. be conscious of) a first-order state, and one of these criteria already concerns indexicality, relocating (or expanding) this criterion to first-order states may not be a significant problem.

The second option for grounding indexicality that appear in line with the HOTTC, is if the grounding is the individual as a whole. This would consist in the idea that the disposition to identify the token of the previous thought is really not a disposition of the HOT, but instead a disposition of the individual⁸. This line of reply seems compatible with Rosenthal’s view. For instance, in the passage quoted above, Rosenthal says “[...] the *individual* is disposed to [...]” (my emphasis). One may now question how we should conceive of the essential indexicality of HOTs if the disposition is a property of the individual and not the state. There are likely issues that need ironing out in this regard, but one way to think about it would that the HOT confers or enables the disposition, rather than is the host to it.

problems and caveats wherever I identified them, I feel certain that readers may discover further advantages, constraints, problems or caveats.

⁸ Note that this has similarity with a view widely held by proponents of the HOTTC’s; namely that ontologically speaking consciousness is a property of *individuals* not a property of states (e.g. Berger, 2014).

Turning to some recent work that may provide an avenue for grounding indexicality within the HOTTC framework, I will consider first the concept of egocentric indexes (EIs) proposed by Tyler Burge (2019). Burge conceives of EIs as the most basic of representational powers, and therefore as essential features for beings with representational minds (2019, p. 49). Initially, this sits well with the HOTTC, given that it is – at its core – a representational theory of consciousness. According to Burge, EIs realize two constitutive functions, the first of which is what is most relevant in the current context, namely that EIs establish the anchor of a representational framework in a contextual and indexical way. This is exactly what the HOTTC needs. Now, Burge’s account of EIs is both lengthy and complex so (depending on which features are adopted by the HOTTC) significant accommodation may be needed. Nevertheless, given that EIs are conceived of as essential powers of representational minds and anchors for indexicality, they provide an enticing platform for establishing a ground for the indexicality of HOTs. Furthermore, because EIs are features of representational minds (or possibly cognitive systems, depending on the details) they can presumably be cashed out in ways not essentially dependant on consciousness, thus avoiding the circularity mentioned above.

The second piece of recent work I will consider here, is the Self-Organizing Metarepresentational Account (SOMA), put forward by Axel Cleeremans and colleagues (2020). SOMA combines a range of claims from existing theories including the radical plasticity thesis (Cleeremans, 2011), and aspects of the HOTTC. According to SOMA individuals develop a model of agenthood by observing and interacting with others in early childhood. The model of agenthood is scaffolded by predictive processing involved in a perception-action loop (that is developed in infancy) and is built up from smaller models of unobservable internal states inferred in – or attributed to – other agents. Once the individual has acquired a model of agenthood, it is available for application to the individual herself. While elements from other theories (and especially from the HOTTC) are clearly recognizable, Cleeremans *et al.* develop SOMA into a theory in its own right. For the present purposes, the details of SOMA are of less importance than the central idea regarding models of agenthood, and the application of these to oneself. According to Cleeremans *et al.* neither the perception-action loop, nor the models of agenthood depend on consciousness (in fact, according to Cleeremans and colleagues, the perception-action loop and the self-other loop – which is the foundation for the models of agenthood – are actually what

result in the ontogenetic development of phenomenal consciousness), thus avoiding the aforementioned circularity. Like with the previously mentioned possibilities, there will be kinks to work out, but *prima facie*, the application of an agenthood-model to oneself, appears to be a good candidate, not only for a grounding of indexicality, but possibly even as the genesis of indexicality.

5. Conclusion

The main aim of this paper has been to show that the regress argument, as presented in Sebastián (2018) does not pose a threat to the HOTTC. I did this by pointing out that Sebastián cannot establish an infinite vicious regress by treating the dispositions of inexistent states on par with actual dispositions of occurrent states. Nevertheless, Sebastián's article sheds light on an aspect of the HOTTC that has not received a lot of attention, namely the indexicality criterion. Thus, while I think the criticism based in the regress argument put forward by Sebastián is ultimately unsuccessful, the paper may prompt additional debate of the indexicality criterion, an upshot I find valuable and interesting. Therefore, to precipitate this further debate into the concept of indexicality within the HOTTC, I have offered up one novel way of thinking about it, *wiz* that the indexicality of HOTs may be transitive. To flesh out this idea, I proposed four distinct possible groundings of the transitive indexicality of HOTs. The proposal of these different grounding possibilities serves a dual purpose. The first is to show that thinking of the indexicality of HOTs as transitive is a viable theoretical option. The second is to provide avenues for future research by delineating theoretical options available for further exploration. I have highlighted that some of these options may not be available to every strand of higher-order theory, which in turn – to my mind – is a good indicator that there may be interesting consequences for, not only the overall HOTTC framework, but also for debates within this framework about the merits of the different theories collected under the HOTTC moniker. I acknowledge in advance that each of the four possible groundings for indexicality provided here – upon further scrutiny – may ultimately be shown to not be a viable option. However, even if this turns out to be the case, additional light has been shed on the indexicality criterion of the HOTTC, a development I would view as positive and wholeheartedly endorse, and I feel certain Sebastián would as well.

Acknowledgements

I wish to express my gratitude, especially to Sebastián for helpful and enjoyable conversation at ASSC19 in addition to subsequent correspondence to clarify his views. I am also grateful to three anonymous reviewers on Filosofiska Notiser, as well as the editor, who has had a not insignificant amount of patience with me.

References

- Berger, J. (2014). Consciousness is not a property of states: A reply to Wilberg. *Philosophical Psychology*, 27(6), pp. 829–842.
- Brown, R. (2015). The HOROR theory of phenomenal consciousness. *Philosophical Studies*, 172(7), pp. 1783–1794.
- Burge, T. (2019). Psychological Content and Egocentric Indexes. In A. Pautz & D. Stoljar (Eds.), *Blockheads!: Essays on Ned Block's Philosophy of Mind and Consciousness* (pp. 41–69): MIT Press.
- Carruthers, P. (2003). Phenomenal consciousness: A naturalistic theory. Cambridge University Press.
- Cleeremans, A. (2011). The Radical Plasticity Thesis: How the Brain Learns to be Conscious. *Frontiers in Psychology*, 2(86).
- Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J.-R., Muñoz-Moldes, S., Vuillaume, L., De Heering, A. (2020). Learning to be conscious. *Trends in cognitive sciences*, 24(2), pp. 112–123.
- Coleman, S. J. (2015). Quotational higher-order thought theory. *Philosophical Studies*, 172(10), pp. 2705–2733.
- Gennaro, R. J. (2006). Between Pure Self-Referentialism and the (Extrinsic) HOT Theory of Consciousness. In U. a. W. Kriegel, Ken (Ed.), *Self-representational approaches to consciousness*: MIT Press.
- Lau, H. (2019). Consciousness, Metacognition, & Perceptual Reality Monitoring. <https://psyarxiv.com/ckbyf/>
- Mandik, P. (2009). Beware of the Unicorn Consciousness as Being Represented and Other Things that Don't Exist. *Journal of Consciousness Studies*, 16(1), pp. 5–36.
- Matey, J. (2006). Two HOTs to handle: The concept of state consciousness in the higher-order thought theory of consciousness. *Philosophical Psychology*, 19(2), pp. 151–175.
- Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, 49(3), pp. 329–359.
- Rosenthal, D. M. (1997). A Theory of Consciousness. In N. Block, O.

- Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 729–753): MIT Press.
- Rosenthal, D. M. (2002a). Explaining Consciousness. In D. Chalmers (Ed.), *Philosophy of Mind: Classical and contemporary readings* (pp. 406–421): Oxford University Press.
- Rosenthal, D. M. (2002b). How many kinds of consciousness? *Consciousness and Cognition*, 11(4), pp. 653–665.
- Rosenthal, D. M. (2004). Being conscious of ourselves. *Monist*, 87(2), pp. 159–181.
- Rosenthal, D. M. (2011). Awareness and Identification of Self. In J. Liu & J. Perry (Eds.), *Consciousness and the Self: New Essays* (pp. 22–50). Cambridge University Press.
- Sebastián, M. Á. (2018). Drop it like it's HOT: a vicious regress for higher-order thought theories. *Philosophical Studies*, 176(6), pp. 1563–1572.
- Weisberg, J. (1999). Active, thin and hot! An actualist response to Carruthers' dispositionalist HOT view. *Psyche*, 5(6), pp. 1–7.
- Weisberg, J. (2010). Misrepresenting Consciousness. *Philosophical Studies*, 154(3), pp. 409–433.

Asger Kirkeby-Hinrup
Cognitive Neuroscience Research Unit
Center for Functionally Integrative Neuroscience,
Aarhus University, Denmark
Theoretical Philosophy
Dept. Philosophy and Cognitive Science, Lund University, Sweden
asger.kirkeby-hinrup@fil.lu.se
akh@cfin.au.dk