

Vicious Dispositions Back to the problem of essential indexicality for Higher-Order Theories

Miguel Ángel Sebastián

Abstract

The *Vicious Dispositions* argument (Sebastián, 2018a) argues that Rosenthal's theory of self-reference—essential for his theory of consciousness—is inadequate because it is implausible that we have the cognitive capacities that it demands. Kirkeby-Hinrup (2020) replies that the argument fails because it assumes that the properties associated with non-existing states have the same ontological status as the properties of existing states. He also suggests alternatives to Rosenthal's account of self-representation.

In this paper I argue that the vicious dispositions argument does not rest on any ontological assumption regarding the properties of non-existing objects and that it only relies on the dispositional characterization of the required capacity to deploy the self concept. I also cast doubts on Kirkeby-Hinrup's positive proposal. My aim is not to argue that it leads to a dead end, but rather contribute to clarifying what the theory requires.

1. Introduction

According to Higher-Order theories, a mental state *M* is conscious just in case one is aware of oneself as being in *M*. Consciousness requires a higher-order mental state that represents *oneself* as being in the conscious state. According to Higher-Order Thought theories the higher-order mental state is a thought. And if we express the content of such a higher-order thought (HOT) in a sentence, we would naturally deploy the first-person pronoun.

Rosenthal (2011, p. 28) stresses that for a mental state to be conscious, it will not suffice that one is aware that someone who happens to be oneself is in that state, one has to be aware of *oneself, as such*. If I am aware that the person (Sebastián) or the entity I am identical with (*X*) is in *M*, that would not result in *M* being conscious unless I am aware that I am Sebastián or that

I am X, for otherwise the thought would not make me aware in any relevant way of myself as being in M. The self-concept seems to refer in an essential indexical way, higher-order representation is *de se* (Castañeda 1966, Chisholm 1981, Lewis 1979, Perry 1979). According to Rosenthal, the subject has to be aware that the subject who has the HOT is the subject the first-person concept deployed in the HOT refers to. But this poses a problem for the theory. If the HOT were to make explicit that the individual who has the HOT is in M, then she would be aware not only of M but also of the HOT, and in this case the HOT would become conscious. But we are not typically conscious of the HOTs. To deal with this problem, Rosenthal proposes that the HOT only identifies tacitly M as belonging to the individual that has the HOT, without the HOT having such explicit content. The remaining challenge for HOT theorists is to explain how the first-person concept refers in a way that is compatible with the theory.

Rosenthal (2011) argues that this is possible if the HOT represents the conscious state belonging to some particular individual S and S is disposed to identify the individual the HOT refers to as the individual that has the HOT. This way, although the HOT does not represent the individual as the thinker of the HOT, the individual is disposed to make such an identification if the question arises.

I have argued in Sebastián (2018a) that Rosenthal's theory of self-reference is inadequate because it is implausible that even we, human adults, have the cognitive capacities that it demands. The reason is that if one is to identify the individual the HOT refers to as the individual that has the HOT, then it has to be possible that the HOT becomes conscious so that the subject knows what is the thought whose subject she has to identify. But this would require that it is possible that the subject has an unconscious third-order thought which also deploys the first-person concept. For this concept to refer in the required way the subject must have the disposition to identify the individual the third-order HOT refers to as the individual that has the third-order thought. But if the subject is to have such a disposition then it has to be possible that the third-order thought becomes conscious... and so on, *ad infinitum*. So, the reference fixing mechanism of the self-concept deployed in the HOT depends upon an arbitrary tall hierarchy of dispositions to have HOTs. And it is implausible that cognitively limited systems like us have such a capacity. Hence, we do not have the explanation of self-representation required by the theory. In what follows, I will refer to this argument as the 'vicious dispositions argument'.

In a recent paper published in this journal, Kirkeby-Hinrup (2020) argues that the vicious dispositions argument fails because it assumes that the properties associated with non-existing states have the same ontological status as the properties of existing states. And he thinks that we cannot derive the conclusion of the argument if they do not. In the second part of the paper, Kirkeby-Hinrup suggests a different alternative to Rosenthal's account of self-representation. I will first show, *pace* Kirkeby-Hinrup, that the vicious dispositions argument does not rest on any ontological assumption regarding the properties of non-existing objects and that it merely rest on the dispositional characterization of the capacity that, according to Rosenthal, gives rise the required first-person representation —a capacity required to have conscious experiences. Then I will cast some doubts on Kirkeby-Hinrup's positive proposal. My aim is not to argue that it leads to a dead end, but rather contribute to clarifying what the theory requires. I applaud the suggestion as I do think that understanding self-reference is the main challenge for higher-order theories and for understanding consciousness in general (Sebastián 2020, MS).

2. The Vicious Dispositions Argument and Inexistent HOTs

The vicious dispositions argument rests on the idea that the capacity to deploy a first-person pronoun in a HOT depends on the capacity to make the HOT conscious (the HOT is typically unconscious) so that the subject can make the required identification. Kirkeby-Hinrup concedes that part of the argument, and I take this possibility to become conscious to be the dispositions of HOTs he refers to. But the HOT can become conscious just in case the subject has the capacity to entertain a 2nd-level HOT, which also has to deploy the first-person pronoun. So it has to be possible that the 2nd-level HOT becomes conscious. Kirkeby-Hinrup also concedes that part, but he thinks that it is a mistake to treat as equal the disposition to become conscious of the HOT, which actually exist, and the disposition to become conscious of the 2nd-level HOT, which is a merely possible state. He thinks that the problem then is analogous to the problem of the classical regress argument against higher-order theories (p. 50).

I am going to show that sharply distinguishing the properties of existing and mere possible states does not help to block the vicious dispositions argument, because the argument only appeals to the capacities that are required from the systems to undergo a conscious experience.

Higher-Order theories endorse what Rosenthal (1997) calls 'the transitivity principle' (Rosenthal and Weisberg 2008, Weisberg 2011), the idea that a state's being conscious consists in one's being conscious of being in that state. The transitivity principle seems to lead to a regress.¹ If for a state to become conscious we have to be conscious of the state, then we need to explain this latter act of consciousness. This in turn requires a third act of consciousness and so on *ad infinitum*. In reply, higher-order theories unpack 'being conscious of' as some sort of higher-order representation, a thought in the case of Rosenthal's theory. And they claim that such a HOT is typically an unconscious one. So, for a state M to become conscious what is required is an unconscious HOT to the effect that one is in M. The regress is stopped because there is no need to appeal to any further conscious state to explain how M becomes conscious. Sure, it might be that the HOT becomes conscious if one has a 2nd-level HOT to the effect that one is in HOT. But the capacity to form 2nd-level HOTs is not necessary for consciousness, only to become conscious of the HOT. Depending on the limits of our cognitive capacities we can only nest a limited number of HOTs, but this is no problem for the theory. Even organisms unable to have 2nd-level HOTs can have conscious experiences: what they cannot have is conscious HOTs. So, according to Rosenthal, there is nothing to fear from the classical regress argument.

We can now turn into the vicious dispositions argument. Rosenthal attempts to explain the capacity that some organisms have for having conscious experiences in terms of their capacity to have a specific kind of thoughts with the content I am in M. This in turn requires an explanation of the capacities required to entertain first-person thoughts: only systems that can entertain first-person thoughts can have conscious experiences. Rosenthal characterizes such a capacity in dispositional terms. The system needs to have the disposition to identify the individual the HOT refers to as the individual that has the HOT. But only systems that can have conscious HOTs can come to make the required identification. So, only systems that can have 2nd-level HOTs have the required disposition. As we have seen, in response to the classical regress argument, Rosenthal argues that there is no need to be able to have 2nd-level HOTs to have conscious experiences. However, it is a consequence of Rosenthal's theory of first-person reference, that only systems capable of 2nd-level HOTs can have conscious experience, because systems that cannot have 2nd-level HOTs cannot have HOTs given that they

1 For detailed discussion see e.g Williford 2006.

cannot have the first-person thoughts required for consciousness. But a system can entertain 2nd-level HOTs just in case it can have thoughts of the form I am in HOT. For the first-person pronoun deployed there to adequately refer, the system has to have the disposition to identify the subject the 2nd-level HOTs refers to as the individual that has the 2nd-level HOTs. But only systems that are able to entertain 3rd-level HOTs have such a capacity. Hence, if a system is not able to have 3rd-level HOTs it does not have the capacity to form HOTs. And so on, *ad infinitum*. So, systems that do not have the capacity to entertain an arbitrarily higher-level HOT do not have the capacity to form 1st-level HOTs because they do not have the disposition to identify the individual the HOT refers to as the individual that has the HOT.

Rosenthal acknowledges that limited cognitive systems like us do not have the capacity to form arbitrarily high-level HOTs. So it would be a problem if the attribution of consciousness would require this capacity. He argues, in reply to the classical regress argument, that his theory does not have such a commitment because the HOT is typically unconscious. However, the vicious dispositions argument shows that Higher-Order Thought theories cannot attribute conscious experiences to systems that cannot entertain arbitrarily high-level HOTs for independent reasons. We have conscious experiences but we lack such a capacity, so the theory cannot be correct. The vicious dispositions argument is not analogous to the classical regress argument and does not rely on thinking that the properties of non-existent entities have the same ontological import as the properties of existing entities: it only depends on the capacities that are required from a system to have the disposition that is supposed to explain first-person representation.

3. Alternative Explanations of the Self Concept

Independently of the merits of Kirkeby-Hinrup's response to the vicious dispositions argument in favor of Higher-order theories, in the second part of the paper he offers an alternative approach for explaining first-person representation. One that is not immune to the vicious disposition argument. I think that the proposal is also problematic.

Kirkeby-Hinrup suggests the possibility that first-person reference in higher-order thoughts derives from a transitive relation. A relation R is transitive if it is such that for any a, b and c, if a is in relation R to b and b is in relation R to c, then a is also in relation R to c. Consider the property of being above. This relation is transitive. If a is above b and b is above c, then c is above a. I am uncertain about the role that transitivity is supposed to play.

The idea seems to be that the n th-level HOT is about oneself in the essential indexical way because it is about the $n-1$ th-level HOT, which is about oneself in the essential indexical way because it is about the $n-2$ th-level HOT, which is about oneself in the essential indexical way because it is about the 1st-level HOT which is about oneself in the essential indexical way. But this does not seem to require that the relation R that holds between the HOTs be a transitive relation. It suffices that it is such that, if the n -level HOT is about oneself in the essential indexical way, and the $n+1$ th-level HOT is in relation R to the n -level HOT then the $n+1$ th-level HOT is also about oneself in the essential indexical way. It seems irrelevant whether if the $n+2$ th-level HOT is in relation R to the $n+1$ th-level HOT, then the $n+2$ th-level HOT is also in relation R to the n th-level HOT —note that at any given time only the conscious n th-level HOT and the $n+1$ higher-level HOT that makes the former conscious exist. What seems to be required is that the higher-level HOT is in relation R to another HOT that is about oneself in the essential indexical way. Be that as it may, the idea, despite being suggestive, faces some problems.

In the first place, we do not have any reason to think that HOTs stand in relation R , independent of offering a response to the vicious dispositions argument. The challenge would be to offer a proper characterization of relation R and then argue that HOTs stand in such a relation. To see the problem, consider that it is natural to think that the relation that HOTs stand in is that of representation. And it is not the case that, in general, a thought about a first-person thought is also about oneself in the essential indexical way. Imagine that you meet your friend Marta and she tells you that she wants to visit her parents. Marta seems to be expressing her first-person thought ‘I want to visit my parents’. And you might form a thought attributing her a first-person thought with the content ‘I want to visit my parents’. The thought that you would entertain in such a situation is about a first-person thought, but it is not itself a first-person thought, what you come to believe is that Marta thinks that she wants to visit her parents. The same would be true if you were the one you are hearing saying ‘I want to visit my parents’ in a recording, but you ignore the fact that it is you the one you are hearing. So, the proposal that HOTs stand in relation R is *ad hoc* and lacks independent motivation. We would need to understand under which conditions thoughts stand in relation R and then evaluate whether HOTs really stand in such a relation.

However, even if we can explain that a thought is about oneself in the essential indexical way because it stands in relation R to another thought that is about oneself in the essential indexical way, we need to explain how the latter comes to be about oneself in the essential indexical way. In the chain of HOTS, how the 1st-level HOT comes to be about oneself in the essential indexical way is left unexplained. We still need a theory of self-representation and we need to explain what it takes for a mental state to be about oneself in the essential indexical way. (Once we have such a theory it is unclear why we need to appeal to relation R to explain how higher-order thoughts are about oneself in the essential indexical way and we could simply appeal to such a theory).

Kirkeby-Hinrup acknowledges this problem and attempts to deal with it toward the end of the paper suggesting that Burge's (2019) Egocentric Indexes or Cleeremans et al.'s (2020) Self-Organizing Metarepresentational Account might do the job.

Egocentric indexes are representational aspects of a psychological state that indexically represent some origin or anchor (a subject in first-person representation) and represent such anchor as carrying systematic, immediate motivational and perspectival psychological consequences for the individual (Burge 2019, p. 49). Kirkeby-Hinrup's suggestion—in line with Burge—is that these representational aspects might ground the acquisition of the first-person concept deployed in the first-person thought—if this is correct there might not be any need to appeal to a relation like R. The idea is suggestive, but we would need i) to have a theory that explains what it takes to have a representational aspect of a psychological state to indexically represent oneself—rather than just assume that we have it—and ii) an explanation of how the first-person concept deployed in HOT is acquired on its basis. Moreover, Burge (2019, sec. IV) denies that egocentric indexes in particular and psychological states refer in a *de se* way against Lewis or Perry. Elsewhere (Sebastián 2020), I have argued that higher-order theories do indeed require *de se* representation of a particular kind—one that gives up the idea that thoughts have truth conditions that do not change from individual to individual—if we want to leave room for the possibility of different individuals having experiences with the same phenomenal character. If my argument there is solid then Burge's Egocentric Indexes cannot play the required role.²

2 For independent criticism of Burge's theory of perception see Echeverri (2017).

Kirkeby-Hinrup also appeals to Cleeremans et al.'s Self-organizing Metarepresentational Account (SOMA). According to this theory, consciousness roughly depends on higher-order representations of first-order internal states, involving the brain redescribing its own representations to itself predicting the influence between the patterns of neural activity in different regions. SOMA is a form of higher-order representational theory, but I fail to appreciate any first-personal component in the theory. It is true that the brain B represents B's representations, but I fail to appreciate any form of essential indexical reference to the brain in the theory. As we have seen in the introduction, Rosenthal requires such a form of representation because if B is aware that B is in M that would not result in any state being conscious unless B also knows that it is B, for otherwise the representation would not make it aware in any relevant way of itself as being in M.

It is true that, in SOMA, a notion of self-awareness plays an important role. But this is not the self-awareness associated with *de se* or first-order representation. Cleeremans and colleagues refer with 'self-awareness' to the sense that we have of being a conscious agent distinct from the world and from other agents (Cleeremans et al. 2020, p. 28). Therefore, such self-awareness seems to depend on consciousness and hence it is not the sort of self-awareness that can ground consciousness —unlike self-awareness understood as awareness of oneself in an essential indexical way does in Higher-Order Thought theory. Interesting as Cleeremans et al. version of a higher-order representational theory might be, I fail to appreciate how it might help to ground the explanation of *de se* representation that Rosenthal's Higher-Order Thought theories requires. It rather seems a rival theory that shares with Higher-Order Thought theory some motivations and empirical predictions.

4. Conclusion

In this paper I have replied to Kirkeby-Hinrup clarifying that the vicious disposition argument presented in Sebastián (2018a) does not rest on the assumption that properties of existing and non-existing objects have the same ontological weight. It merely rests on the capacities required from a system to have the dispositional property that according to Rosenthal grounds essential indexical reference and consciousness with it. Conscious beings should have such a capacity but we, conscious beings, do not have it. As a consequence the theory has to be abandoned.

I have also reviewed Kirkeby-Hinrup's alternative suggestions to ground

the required essential indexical reference of the self-concept. Although I have argued that they are not ultimately adequate, I think that this line of research is definitely the one that has to be pursued if the idea that consciousness is to be grounded on higher-order thoughts is to get off the ground. I myself have attempted to deal with this problem elsewhere (e.g. Sebastián 2012, 2018b).³

References

- Burge, T. (2019). Psychological content and egocentric indexes. In A. Pautz and D. Stoljar (eds), *Blockheads!: Essays on Ned Block's philosophy of mind and consciousness*, MIT Press.
- Castañeda, H.-N. (1966). 'He': A study in the logic of self-consciousness. *Ratio* 8, pp. 130–157.
- Chisholm, R. M. (1981). *The First Person: An Essay on Reference and Intentionality*. Minneapolis: University of Minnesota Press.
- Cleeremans, A., Achoui, D., A. Beauny, Keuninckx, L., Martin, J., Moldes, S. M., Vuillaume, L. and de Heering, A. (2020). Learning to be conscious. *Trends in Cognitive Sciences* 24(2), pp. 112– 123.
- Echeverri, S. (2017). Visual reference and iconic content. *Philosophy of Science* 84, pp. 761–781.
- Kirkeby-Hinrup, A. (2020). Essentially Indexical Higher-Order Thoughts and Dispositions of Inexistent Mental States. *Filosofiska Notiser* 7(2), pp. 45–59.
- Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review* 88(4), pp. 513–543.
- Perry, J. (1979). The problem of the essential indexical. *Noûs* 13, pp. 3–21.
- Rosenthal, D. (1997). A theory of consciousness. In N. Block, O. J. Flanagan and G. Guzeldere (eds), *The Nature of Consciousness*, MIT Press.
- Rosenthal, D. (2011). Awareness and identification of self. In J. Liu and J. Perry (eds), *Consciousness and the Self: New Essays*, Cambridge: Cambridge University Press.
- Rosenthal, D. and Weisberg, J. (2008). Higher-order theories of consciousness. *Scholarpedia*, 3(5), 4407.
- Sebastián, M. Á. (2012). Experiential awareness: Do you prefer "it" to "me"? *Philosophical Topics* 40(2), pp. 155–177.

3 Financial support for this research was provided by DGAPA projects IA400218 and IG400219, and the Spanish Ministry of Science, Innovation and Universities via research Grant PGC2018-095909-B-100.

Miguel Ángel Sebastián

- Sebastián, M. Á. (2018a). Drop it like it's HOT: a vicious regress for higher-order thought theories. *Philosophical Studies*, 176(6), pp. 1563–1572.
- Sebastián, M. Á. (2018b). Embodied appearance properties and subjectivity. *Adaptive Behaviour* 26(5), pp. 199–210.
- Sebastián, M. Á. (2020). Being self-involved without thinking about it: Confusions, virtues and challenges of higher-order theories. In J. Weisberg (ed.), *Qualitative Consciousness: Themes from the Philosophy of David Rosenthal*, Cambridge: Cambridge University Press.
- Sebastián, M. Á. (MS). First-person perspective in experience: Core de se representations as an explanation of the subjective character of experience. *Unpublished*.
- Weisberg, J. (2011). Misrepresenting consciousness. *Philosophical studies* 154(3), pp. 409–433.
- Williford, K. (2006). The self-representational structure of consciousness. In U. Kriegel and K. Williford (eds), *Self-Representational Approaches to Consciousness*, The MIT Press.

Dr. Miguel Ángel Sebastián
msebastian@gmail.com
www.scc-philab.com/sebastian
Instituto de Investigaciones Filosóficas, UNAM