

Drop it like it's HOT: a vicious regress for higher-order thought theories

Miguel Ángel Sebastián

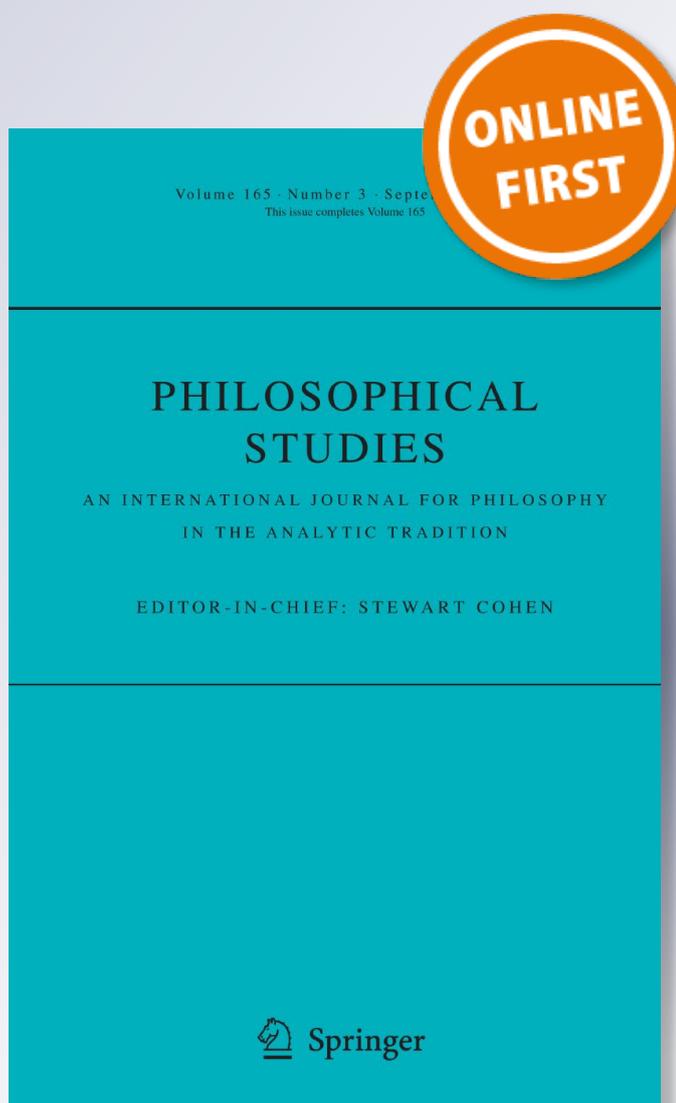
Philosophical Studies

An International Journal for Philosophy
in the Analytic Tradition

ISSN 0031-8116

Philos Stud

DOI 10.1007/s11098-018-1078-7



Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media B.V., part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Drop it like it's HOT: a vicious regress for higher-order thought theories

Miguel Ángel Sebastián¹ 

© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract Higher-order thought (HOT) theories of consciousness attempt to explain what it takes for a mental state to be conscious, rather than unconscious, by means of a HOT that represents oneself as being in the state in question. Rosenthal (in: Liu, Perry (eds) *Consciousness and the self: new essays*, Cambridge University Press, Cambridge, 2011) stresses that the way we are aware of our own conscious states requires essentially indexical self-reference. The challenge for defenders of HOT theories is to show that there is a way to explain the required reference-fixing mechanisms that is compatible with the theory. According to Rosenthal, the reference to oneself as such is grounded in the disposition to identify the individual the HOT refers to as the individual who has that HOT. I argue that this leads to a vicious infinite regress on the more than plausible assumption that our cognitive capacities are limited. This leaves such theories without a foundation, since self-reference is thought essential to consciousness.

Keywords Consciousness · Higher-order theories · Essential indexical · Transitivity principle · Subjective character

1 Introduction

Higher-order theories of consciousness attempt to explain what it takes for a mental state to be conscious, rather than unconscious, by means of some higher-order awareness. According to this family of theories, in having an experience, say, as of a red apple, one is not merely aware of the red apple but also aware of oneself as

✉ Miguel Ángel Sebastián
msebastian@gmail.com

¹ Instituto de Investigaciones Filosóficas, Universidad Autónoma de México, UNAM, Circuito Mtro Mario de la Cueva, Ciudad Universitaria, Del. Coyoacán, 04510 Mexico, D.F., Mexico

being in a certain state. The rationale for the claim that such higher-order awareness is required for consciousness is that states we are completely unaware of would not be conscious.

Higher-order theorists disagree with regard to the best way to characterize the necessary higher-order awareness. The main dispute is whether such awareness is perception-like or thought-like. Theories that maintain that the latter is the case are called higher-order thought (HOT) theories (Brown 2015; Gennaro 1996, 2012; Rosenthal 1997, 2005); while theories that defend the former as being the case are known as higher-order perception (HOP) or 'inner-sense' theories (Armstrong 1968; Carruthers 2000; Lycan 1996). According to HOT theories, in normal conditions, when I have a phenomenally conscious experience as of red, I am in a mental state with certain content, let us call this content 'red*'. For this mental state to be phenomenally conscious, there has additionally to be a HOT targeting it, whose content is something like 'I am seeing red*'. In contrast, HOP theories maintain that what is required is a (quasi-)perceptual state directed at the first-order state.¹

Independently of the disagreement, higher-order theories can be characterized by their endorsement of what Rosenthal (1997) calls 'the transitivity principle' (Rosenthal and Weisberg 2008): the idea that a state's being conscious consists in one's being conscious of that state. The whiff of regress is a well-known problem that can be traced back to Aristotle (Caston 2002): if a state's being conscious requires one's being conscious of that state, and this in turn another conscious state, then we fall into a vicious regress (for detailed discussion see Kriegel 2009b, ch. 4). In reply, defenders of higher-order theories deny that the higher-order state that makes us conscious of the first-order mental state has to be a conscious one, thereby avoiding the regress.²

In this paper, I focus on HOT theories. I will show that, independently of the foregoing reasoning, such theories do indeed fall prey to a vicious regress due to the fixing mechanism of the first-person concept deployed in the HOT.

2 HOT theories and the essential indexical

According to HOT theories, a mental state is conscious in virtue of the right kind—assertoric and seemingly non-inferential—of HOT to the effect that *I* am in a certain state. For example, when one is having an experience as of red, one would entertain a HOT to the effect that one herself is seeing red or that one herself is visually representing red: the intentional content of the HOT is something like *I, myself, am representing red* (Brown 2015; Lau and Rosenthal 2011). Consequently, HOT

¹ A second point of disagreement is whether a given state is conscious in virtue of its actually being the target of a higher-order representation (Brown 2015; Rosenthal 1997, 2005) or merely by the disposition towards such higher-order representation (Carruthers 2000).

² This might lead to the further problem of being unable to provide a justification for the transitivity principle (Kriegel 2009a). An alternative solution is offered by self-representational theorists (Caston 2002; Kriegel 2009b; Sebastián 2012), who might accept something along the lines of the transitivity principle while denying that consciousness is something that could be rendered by a different state.

theorists agree on the idea that the kind of higher-order thought that renders a mental state conscious deploys the first-person concept. The objection to HOT theories that I am presenting in this paper arises precisely from the reference-fixing mechanism of such a concept.

The need for the deployment of the first-person concept might sound several alarms, as many have doubted that human babies and non-human animals have such a concept, while we have few reasons to doubt that they can have conscious experiences. In reply to those worries, Gennaro (2012) has claimed that there are different “I-concepts” that might play the role required by HOT theories: (1) I *qua* this thing (or body); (2) I *qua* experiencer of mental states; (3) I *qua* enduring thinking thing; and (4) I *qua* thinker among other thinkers. Gennaro also discusses a range of empirical evidence and convincingly shows that babies, as well as most animals, might at least have an I-concept given by (1). However, it is not clear that such a concept would suffice for the theory; something closer to (2) seems to be required. As Rosenthal (2011, p. 28) stresses, the way we are aware of our own conscious states requires essentially indexical self-reference (Perry 1979; see also Castañeda 1966; Chisholm 1981; Lewis 1979) and hence not any “I-concept” can do the job. The reason is that for a mental state to be conscious, it does not suffice that one is aware that someone who happens to be oneself is in that state—that is, even assuming that one is identical with a certain body, being aware that this body is in a certain state (as per (1)) will not do the job—rather, one has to be aware of oneself, as such, as being in the state. Following Rosenthal’s example, if I were identical to this body and aware that this body is in pain, that would not result in any pain being conscious unless I were aware that I am this body; for otherwise, the thought would not make me aware in any relevant way of myself as being in pain.

Once it is clear that the concept deployed in HOT theories refers in an essentially indexical way, I do not want to press any further the controversial conceptual capacities of babies and non-human animals (although see fn. 4 below): I prefer to focus on the capacities of typical human adults. The challenge for defenders of HOT theories is to show that there is a way to explain the required reference-fixing mechanisms that is compatible with those theories given the capacities that we have (Zahavi and Parnas 1998). For example, Recanati’s proposal (2007, 2012) won’t do it because, according to his theory, the reference of the first-person concept is grounded in the experience itself (see also Peacocke 2014), whereas HOT theories are committed to the claim that the first-person concept is more fundamental than conscious experience.

Despite the relevance of this question to the evaluation of the coherence of the view, it has received very little attention, with Rosenthal’s (2011) proposal being an honorable exception. In the next section I present his view and then argue that it leads to a vicious regress, thereby leaving HOT theories with no foundation.

3 A vicious regress

According to Rosenthal, the HOT that I, myself, am in mental state M, represents as being in mental state M the very same individual who thinks that HOT. The required self-reference is grounded in the tacit identification of oneself as the thinker of the HOT: one refers to oneself as such, when one refers to something as the individual doing the referring. So, if I am having a conscious pain, then following the transitivity principle I am aware of myself as having that pain. According to HOT theories, this in turn requires a HOT to the effect that I myself have that pain. This HOT refers in the required essentially indexical way, by means of attributing the pain to the very same individual who has that HOT.

Rosenthal notes a complication for HOT theories that is derived from the fact that HOTs are typically unconscious thoughts and they only become conscious when we are aware of having them; i.e., when we have a third-order thought about them.³ According to the theory, a state, M, is conscious just in case one is aware of oneself as being in M. Such an awareness is explained by the deployment of an unconscious thought to the effect that oneself as such (as the individual that is having the thought) is in M. So, being the target of a HOT suffices for consciousness, and this applies to HOTs too. Now, if the HOT is to be unconscious, then it should not make it explicit that the individual that has the HOT is in M, for this would make the individual that is having the HOT aware of having the HOT, and hence conscious. As Rosenthal stresses:

Consider the case of conscious pain. If a HOT had the explicit content that the individual who thinks this thought is in pain, that would make one aware not just of the pain, but of the HOT itself. And since HOTs are seldom conscious, they cannot have explicit content of that sort. So it cannot be that the HOT explicitly identifies the individual that is in pain as the thinker of that HOT; rather the HOT identifies the individual in that way tacitly, only in effect. (Rosenthal 2011, p. 30)

Hence, the HOT only tacitly identifies the pain as belonging to the individual who has the HOT, without the HOT having any such explicit content. Now, an explanation of such tacit identification is required. And Rosenthal provides it. The conjunction of the following two elements explains, according to Rosenthal, how the first-person concept refers in the required essentially indexical way:

1. The HOT represents the pain as belonging to some particular individual S.
2. S is disposed to identify the individual the HOT refers to as the individual who has that HOT.

In 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 should the question arise—of course, as Rosenthal stresses, such a question seldom arises and thus the

³ Recall from the introduction that HOTs are typically not conscious. This is required if we want to avoid the regress to which the transitivity principle would otherwise apparently lead us.

Drop it like it's HOT: a vicious regress for higher-order...

individual might never actually perform the identification in question.⁴ This disposition constitutes the tacit identification of the self required by the HOT, and with it its essentially indexical self-reference (2011, p. 30). Unfortunately, it is also this disposition that gives rise to a vicious infinite regress.

4 Why there is an infinite regress

The problem with Rosenthal's proposal is derived from the fact that if one is to identify the individual the HOT refers to as the individual who has the HOT, then it has to be cognitively possible for the HOT to become conscious. Let me pause to clarify and justify this claim.

Dispositional properties seem to be related to certain possibilities concerning the entity that instantiates the disposition, rather than with the actual behavior of the entity. For this reason—and despite not being rare in the philosophy of mind and sometimes, as in Rosenthal's case, very illuminating—explanations of occurrent properties (such as referring, in the way the first-person concept deployed in the HOT refers) in terms of dispositional ones seem to be wanting. If the explanation in question is to be in any sense satisfactory, at least to a first approximation, the disposition has to be a canonical one (Lewis 1997): they have to be explicit about their manifestation and stimulus conditions (the conditions under which the manifestation is triggered). Thus, if the stimulus conditions are not provided, we lack the tools to assess the theory.

In Rosenthal's theory, the manifestation is an action that the subject performs: the identification of the individual the HOT refers to as the individual who has that HOT. On top of that, we need the stimulus conditions; and whatever those stimulus conditions might be, they would require the HOT to be a conscious one if the subject is to identify the individual that the HOT refers to. If this were not the case, the subject would be unable to recognize the HOT concerning which the required action is to be performed.⁵

⁴ Even if one concedes that non-human animals and human babies have the conceptual abilities required by HOT, it is unreasonable to assume that they are able to establish the identification in question. In reply, Rosenthal claims that they:

[H]ave no irrelevant, inessential ways of referring to themselves in thought. They do distinguish themselves from everything else, and can thereby refer to themselves in thought. But their HOTs do not require the essential indexical, since distinguishing themselves from everything else provides the only way they have to refer to themselves (ibid. p. 34).

This seems to be a gratuitous claim and we have good reasons to think that it is false. The capacity to attribute mental states to others has been demonstrated in, for example, corvids and canids (Bugnyar and Heinrich 2006; Hare and Tomasello 2005; Stulp et al. 2009; Udell et al. 2008). One would expect that by seeing their own image reflected in, say, a river, they might attribute a certain mental state to the animal they are seeing. Therefore, it seems false that they lack inessential ways to refer to themselves and hence the problem of fixing the reference of the concept in their HOTs persists.

⁵ What is more, as an anonymous referee has remarked to me, in order to report that the individual the HOT refers to is the same as the one who has the HOT seems to require thinking about the HOT. This in turn would make the HOT conscious according to the theory.

One might try to resist this conclusion by insisting that the identification in question could be performed at an unconscious level. However, this does not seem to be an option because we have no reason whatsoever to think that we have such a disposition. Whereas Rosenthal gives us reasons to think that we can make the identification if the HOT is a conscious one (ibid. fn. 11), we have no reason independent of the theory to think that we could have an unconscious manifestation of the disposition. Yet, this is precisely what we require if the theory is to get off the ground. Moreover, it is unclear in what sense, if the identification were an unconscious one, it would be an identification we could attribute to the subject—rather than to a subsystem thereof—as the theory demands.

Fortunately, Rosenthal is clear regarding the stimulus condition, which would be a question concerning who the individual who has the HOT is: “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.*” (ibid. p. 30, my emphasis). But Rosenthal is implicitly assuming that the HOT is a conscious one here. As he stresses, “the question seldom if ever does arise; so the individual that has the HOT may never actually perform that identification” (ibid. p. 30) and hence the HOT can typically be unconscious, as the theory requires (for, as we have seen, the explicit identification would render the HOT conscious). If the HOT were unconscious, in reply to the question the subject would deny any knowledge of what thought the person asking the question was talking about. The subject would not make any identification, and the theory would be falsified, because the stimulus condition would obtain but the manifestation would not be triggered.

So, 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; a third-order thought whose intentional content is something like *I, myself, am in a state that represents that I, myself, am in a state that represents red.* 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 that deploys the first-person concept, which refers to oneself in an essentially indexical way, and this thought too must be capable of consciousness, and so on: ad infinitum.

5 Why the regress is vicious

If my reasoning is correct, the reference-fixing mechanism of the self concept deployed in the HOT, and hence, according to the theories we are considering, the capacity to undergo phenomenally conscious experiences, requires an arbitrarily tall

hierarchy of dispositions to have HOTs.⁶ Beings like us have the capacity to have phenomenally conscious experiences, therefore the truth of the theory depends on whether beings like us, in the actual world or a sufficiently close one (in order to evaluate the theory only those worlds including beings with the cognitive capacities we have are relevant), have the disposition in question. And whether or not we have such a disposition is determined by whether or not, in the actual world or in a sufficiently close one, the hierarchical chain of dispositions in question can be manifested. This does not mean that there have to be circumstances under which the subject can entertain all the thoughts, but just that for each HOT there have to be circumstances under which they can be entertained. So, the reference of the first person concept deployed in the n th-order thought requires there to be circumstances under which, in the actual world or in a sufficiently close one, we entertain a thought with the intentional content that *I, myself, am in a state that represents that I, myself, am in a state that represents that... that I, myself, am in a state that represents red.*⁷

If our cognitive capacities are limited, we cannot entertain thoughts such as the previous one of infinite length: there is a maximally demanding HOT we can entertain. Let us call it maximal-order thought (MOT). MOT deploys a self concept that refers in an essentially indexical way; and this is so in virtue of the disposition to identify the individual MOT refers to as the individual who has that MOT. However, this requires that MOT *can* be conscious, which it cannot be because that would require that we entertain a thought about MOT, and *ex-hypothesi*, we cannot. This in turn entails that MOT-1 (the HOT that MOT refers to) cannot be conscious, because the self concept in MOT cannot refer in an essentially indexical way. And hence, the subject lacks the disposition to identify the subject MOT-1 refers to as the individual who has MOT-1, so the self concept in MOT-1 does not refer in an essentially indexical way. And this means that MOT-2 cannot be conscious. Finally, following this chain, the self concept in second-order thought does not refer in an essentially indexical way and hence we cannot be aware of ourselves as having the experience. But according to HOT, this is a necessary condition for having conscious experiences. And we do have conscious experiences, so HOT theory fails on its own terms to give a model for consciousness.

A final remark regarding the possibility of the manifestation is in order. Summarizing the argument so far, we have seen that HOT theories attempt to

⁶ Needless to say, an hypothetical HOT theory that manages to resist the claim that the first-person concept is deployed in the required HOT would not be targeted by my argument—see Coleman (2015).

⁷ It is worth introducing a caveat at this point. This reasoning does not assume that the accessibility relation between worlds is transitive. For the first-person concept deployed in the HOT in w_1 to refer, it is necessary for there to be an accessible world, w_2 , in which the HOT is conscious. This means that a third-order thought—that also deploys the first-person concept—is entertained in w_2 . So, again, this requires that there is a possible world, w_3 , which is accessible from w_2 , where the third-order thought is conscious; and so on, ad infinitum. Now, my argument does not assume that w_3 , nor any posterior world in the chain, has to be accessible from the actual world, w_1 . That would be a mistake, because there is no need for the required accessibility relation to be transitive. It only assumes that the cognitive capacities of beings like us are what is relevant for assessing the theory, and that our cognitive capacities are limited. This means that there is a world, w_n , in which a thought of the maximal length beings like us can entertain is entertained and therefore, that w_{n+1} is not accessible from w_n . I am grateful to Alessandro Torza for pressing me on this point.

explain what distinguishes conscious states from other kind of states: what is required from a cognitive system in order to have a conscious state. They claim that what is required is a HOT to the effect that one herself is in the first-order state. So, having an experience requires entertaining a thought that deploys the first-person concept which refers in a seemingly essentially indexical way. And this in turn requires that one is disposed to make a certain identification. One has an experience as of red, only if one entertains a HOT whose content is something like 'I myself am representing red', and for such first-person concept to refer in the required way, one needs to have the disposition to answer the question about who the 'I' in the HOT refers to. So, according to the theory, if the subject can entertain conscious states, she has the required disposition. Such a disposition entails, in turn, the disposition to form an arbitrarily high higher-order thought as I have shown. But cognitively limited systems like us cannot form such an arbitrarily high higher-order thought and hence lack the required disposition and the capacity to entertain conscious states with it. In reply, one might note that there are cases in which we are willing to accept that an object has a dispositional property despite the fact that the possibility of the manifestation is prevented. A paradigmatic example of such cases are what Johnston (1992) calls 'maskers'. Consider, for example, the dispositional property of being fragile. A fragile glass has the disposition of breaking when struck. However, imagine that the fragile glass is carefully protected by some kind of packaging. In such cases, something external to the object to which the dispositional property is attributed prevents the possibility of the manifestation. However, we would be willing to accept that the object still has the dispositional property [probably because dispositions are intimately related to intrinsic properties (Lewis 1997)]: we would like to say that the glass is still fragile and disposed to break when struck, despite the fact that it will not break when struck because of the work of the packaging. In parallel, it might be the case that the disposition to make some identifications cannot be manifested because it is not possible that anybody will ask the question that would trigger the manifestation of the disposition. However, we would like to maintain that the subject still has the disposition to do so, if someone were to ask. The identification cannot be made in this case, due to something external and unrelated to the capacities of the system on which the capacity of having conscious states depends. But this sense in which the disposition might fail to manifest is different from the one pointed out above, and helps to illustrate why the impossibility of making the identification in question is defeating for the theory.⁸

The argument above does not point to the presence of external maskers but to the limits of the very same cognitive architecture that HOT theories attempt to characterize. The capacity to have conscious states is something that some cognitively limited systems like us have. However, it is unclear in what sense cognitively limited systems have the disposition to entertain an arbitrarily high HOT, which is a prediction of the theory as I have shown. It is not a contingent matter, as in the case of maskers for example, that there is a maximally high HOT

⁸ I am grateful to an anonymous referee for pressing me on this point.

that a cognitively limited system can entertain. Necessarily, a cognitively limited system (the kind of systems we aim to characterize) cannot entertain an infinitely high HOT. But this, as I have shown, is what would be required in order to ground the reference of the first-person concept deployed in the higher-order thought that renders a state conscious in the first place.

6 Conclusion

HOT theories claim that a mental state is conscious when one has a higher-order thought to the effect that oneself is in the corresponding conscious state. This demands an explanation of how the self concept deployed in the HOT refers in the required essentially indexical way that is compatible with the truth of the theory and plausible assumptions about our cognitive architecture.

Rosenthal (2011), to the best of my knowledge, is the only author who has risen to this challenge. He proposes that the self concept refers in virtue of the disposition to identify, should the question arise, the individual the HOT refers to as the individual who has that HOT.

In this paper I have argued that such a disposition entails the possibility that the HOT is conscious, and that this leads to a vicious infinite regress. Therefore the theory fails to provide a satisfactory account of consciousness.

Acknowledgements I am grateful to Paloma Atencia-Linares, Axel Barceló, Denis Buehler, Elvira Di Bona, Manolo Martínez, Ricardo Mena, Angélica Pena-Martínez, James Stazicker, Nils-Hennes Stear (who suggested to me the title for the paper), and Alessandro Torza for discussion and comments on a previous draft. I owe a special debt to David Rosenthal for helpful discussion of the argument presented in this paper and also to an anonymous referee of this journal. A version of this paper was presented at the Congress of the Italian Society for Analytic Philosophy. I am grateful to the participants, especially to Krisztina Orban, Michelle Palmira, Carlota Serranilla and Hong Yu Wong. Financial support for this research was provided by the Spanish Government (FFI2014-51811-P), the British Academy and the PAPIIT (IN400218).

References

- Armstrong, D. (1968). *A materialist theory of the mind*. London: Routledge.
- Brown, R. (2015). The HOROR theory of phenomenal consciousness. *Philosophical Studies*, 172, 1783–1794.
- Bugnyar, T., & Heinrich, B. (2006). Pilfering ravens, *corvus corax*, adjust their behavior to social context and identity of competitors. *Animal Cognition*, 9, 369–376.
- Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge: Cambridge University Press.
- Castañeda, H. N. (1966). 'He': A study in the logic of self-consciousness. *Ratio*, 8, 130–157.
- Caston, V. (2002). Aristotle on consciousness. *Mind*, 111(444), 751–815.
- Chisholm, R. M. (1981). *The first person: An essay on reference and intentionality*. Minneapolis: University of Minnesota Press.
- Coleman, S. (2015). Quotational higher-order thought theory. *Philosophical Studies*, 172, 2705–2733.
- Gennaro, R. (1996). *Consciousness and self-consciousness: A defense of the higher-order thought theory of consciousness*. Amsterdam: John Benjamins.
- Gennaro, R. (2012). *The consciousness paradox: Consciousness, concepts, and higher-order thoughts*. Cambridge: MIT Press.

- Hare, B., & Tomasello, M. (2005). Human-like social skills in dogs? *Trends in Cognitive Sciences*, 9, 438–444.
- Johnston, M. (1992). How to speak of the colors. *Philosophical Studies*, 68, 221–263.
- Kriegel, U. (2009a). Self-representationalism and phenomenology. *Philosophical Studies*, 143(357–381), 2009a.
- Kriegel, U. (2009b). *Subjective consciousness: A self-representational theory*. New York: Oxford University Press.
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15(8), 365–373.
- Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review*, 88(4), 513–543.
- Lewis, D. (1997). Finkish dispositions. *The Philosophical Quarterly*, 47, 143–158.
- Lycan, W. (1996). *Consciousness and experience*. Cambridge: MIT Press.
- Peacocke, C. (2014). *The mirror of the world: Subjects, consciousness and self-consciousness*. Oxford: Oxford University Press.
- Perry, J. (1979). The problem of the essential indexical. *Noûs*, 13, 3–21.
- Recanati, F. (2007). *Perspectival thought: A plea for (moderate) relativism*. Oxford: Oxford University Press.
- Recanati, F. (2012). Immunity to error through misidentification: What it is and where it comes from. In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification: New essays*. Cambridge: Cambridge University Press.
- Rosenthal, D. (1997). A theory of consciousness. In N. Block, O. J. Flanagan, & G. Guzeldere (Eds.), *The nature of consciousness*. Cambridge: MIT Press.
- Rosenthal, D. (2005). *Consciousness and mind*. Oxford: Oxford University Press.
- Rosenthal, D. (2011). Awareness and identification of self. In J. L. Liu & J. Perry (Eds.), *Consciousness and the self: New essays*. Cambridge: Cambridge University Press.
- Rosenthal, D., & Weisberg, J. (2008). Higher-order theories of consciousness. *Scholarpedia*, 3(5), 4407.
- Sebastián, M. (2012). Experiential awareness: Do you prefer it to me? *Philosophical Topics*, 40(2), 155–177.
- Stulp, G., Emery, N., Verhulst, S., & Clayton, N. (2009). Western scrub-jays conceal auditory information when competitors can hear but cannot see. *Biology Letters*, 5, 583–5859.
- Udell, M., Dorey, N., & Wynne, C. (2008). Wolves outperform dogs in following human social cues. *Animal Behavior*, 76, 1767–1773.
- Zahavi, D., & Parnas, J. (1998). Phenomenal consciousness and self-awareness: A phenomenological critique of representational theory. *Journal of Consciousness Studies*, 5, 687–705.