

# Change Blindness and Misrepresentation

**Asger Kirkeby-Hinrup**  
Lund University

BIBLID [0873-626X (2016) 42; pp. 37-56]

## **Abstract**

Some proponents of the higher-order thought (HOT) theory of consciousness defend the view that higher-order misrepresentation is possible. In support of this view they have proposed various pieces of empirical evidence. This paper examines one such piece of proposed empirical evidence; Change blindness (CB). CB occurs when a subject fails to detect salient changes in visual scenes. I propose an alternative interpretation of the CB phenomenon on which misrepresentation does not occur. Finally, I examine three lines of reply that might be pursued to defend the claim that CB is evidence of misrepresentation against my interpretation. I conclude that none of the lines of reply succeed in preserving CB as evidence of misrepresentation. The upshot is that, pending further evidence, CB cannot be deployed as empirical evidence in the debate on misrepresentation.

## **Keywords**

Consciousness, Misrepresentation, Higher-order thought, Change Blindness, HOT.

## 1 Introduction

Higher-order theories of consciousness assume that mental states can be unconscious or conscious. The underlying distinction is that a mental state that one is not aware of being in is not — on any reasonable account — a conscious state. This idea has become known as the *Transitivity Principle* (TP): *A conscious state is a state one is aware of oneself as being in.*

Higher-order theories propose that being conscious of something may be explained in terms of a relation between two mental states. The higher-order thought theory of consciousness suggests a relation obtaining between the conscious state and a thought-like

*Disputatio*, Vol. VIII, No. 42, May 2016

Received: 14/03/2014 Revised: 06/02/2015 Accepted: 22/02/2016

higher-order representation (HOT) of it, is what generates conscious experience of being in a mental state. Thus, the way that one becomes conscious of being in a given mental state is by having a suitable HOT about that state. The HOT makes you conscious of being in a mental state  $x$  because it represents the thinker of the thought as being in  $x$ .

On David Rosenthal's version of HOT theory, that will be the topic below, we become aware of ourselves as being in a given conscious state by having a suitable higher-order thought about that state of the form 'I am in state  $x$ '. There are several riders, as to what qualifies as a *suitable* HOT. The riders include, for example, provisions that the HOT is not itself a conscious state, the HOT must not come about as a result of conscious inference or observation, and that the HOT must be simultaneous with the mental state it represents (see e.g. Rosenthal 1997, 2002a, 2002b for the details of HOT theory). In addition, Rosenthal (e.g. 2004b) thinks that the presence of a suitably caused HOT is sufficient for consciousness, even if the target state of the HOT does not exist. An individual will undergo a conscious experience of red provided that she believes that she is undergoing an experience of red. This is known as the possibility of misrepresentation and has received much attention in the past decade.

Misrepresentation occurs if there are discrepancies between a HOT and the mental state it is about. When there is a discrepancy, an individual will have a conscious experience of the state the HOT describes her as being in — regardless if she has it or not. Thus, if a HOT describes an individual as seeing something blue, the individual will experience seeing blue regardless of which lower-order states obtain<sup>1</sup>. According to Rosenthal, the mental state the individual is conscious of being in, in this case, is merely a "notional state" (e.g. Rosenthal 2000b: 232). A notional state is merely a *de dicto* state, it has no neural correlates; it suffers from intentional inexistence.

It has been argued that if discrepancies between a HOT and the state that it is about are possible, this poses serious problems for the

<sup>1</sup> Observe that some versions of the HOT theory reject the possibility of misrepresentation, e.g. the dispositionalist HOT theory of Peter Carruthers (1998) and the wide intrinsicity view of Rocco Gennaro (1996, 2012).

HOT theory. The two problems that have been discussed most extensively are whether the HOT theory violates intuitive theoretical principles associated with first, phenomenal qualities, and second, existence criteria of mental states. In the first case the underlying idea is that phenomenal qualities are intrinsic to first-order states and do not depend on HOTs (e.g. Balog 2000, Block 2011a, 2011b, Matey 2011, Neander 1998, Rosenthal 2004b, 2011, 2012, Shepherd 2013, Weisberg 2008, 2010, 2011). In the second case, the objection is that if a mental state is conscious this seems to entail that it exists, therefore explaining misrepresentation in terms of a “notional state” cannot be done (e.g. Berger 2014, Block 2011a, 2011b, Matey 2006, Rosenthal 2011, 2012, Weisberg 2010, Wilberg 2010). Consequently, according to the critics, misrepresentation along the lines described above, would be impossible.

While the HOT theory seems counterintuitive to some, it is not incoherent: It can account for its explanandum within the framework it has set forth for itself. Many defendants of the HOT theory consider themselves to have provided successful replies to the theoretical objections against the theory (e.g. Rosenthal 2011, 2012). Of course, the opponents do not agree. There is no obvious winner of the debate, which seems to have reached a stalemate. Now, if it is correct that the HOT theory in general, and the stance that misrepresentation is possible in particular, both are coherent, and, furthermore, that there is limited interest in pursuing the theoretical debate, there seems to be little hope to resolve the issue of misrepresentation by theoretical argumentation. However, there is, in addition to theoretical argument, the option of invoking empirical data in the debate on misrepresentation. One might object that empirical data cannot have any bearing on conceptual issues. While this objection is legitimate, the aim is not to let the empirical data decide the conceptual issue. Rather, the inclusion of empirical data in the debate is a way of showing that the HOT theory has empirical force. Furthermore, it is part of a larger inference to the best explanation (IBE) argument in support of the HOT theory against other theories of consciousness. The IBE argument subsumes the misrepresentation debate. In the context of the IBE argument, defendants of the HOT theory have put forth different suggestions of how empirical data can support both the theory in general and specifically, claims

for the possibility of misrepresentation (Lau 2007, Lau and Brown *to appear*; Lau and Rosenthal 2011a, 2011b, Rosenthal 2012).<sup>2</sup> To evaluate these suggestions amounts to assess the data and their bearing on the issues they concern. So far, not much work has been carried out in this regard. Two recent exceptions include Kirkeby-Hinrup 2014 and Sebastián 2014. This article is a further effort in this direction.

One piece of empirical data that has been suggested to support the possibility of misrepresentation is the phenomenon of *change blindness* (CB). In CB paradigms, limitations in the visual system are exploited to make the subject fail to detect changes between presented visual stimuli that are similar to each other but not identical. E.g., when two images are presented in close temporal proximity experimental manipulations can prevent the subject from detecting salient differences between them. On the interpretation offered by the defendants of the possibility of misrepresentation, the failure to detect changes across scenes can be explained by the subject misrepresenting her visual states. The idea is that when an individual fails to detect the changes in his visual stimuli, when there is good reason to think the corresponding first-order states have changed. Thus, what she is consciously experiencing exhibit discrepancies with the states she is actually in. The aim here is to investigate this particular interpretation of CB. The upshot of the present investigation is the observation that, because correct representation is distinguished from misrepresentation by how the HOT relates to the mental state that it concerns, misrepresentation cannot be separated from correct representation unless the representational structure is known. However, because at present there is no way of knowing which representational structures obtain in the experiments, viz. because there is no way of establishing how the HOT relates to the mental state it is about, the empirical data from CB can have very little bearing on the misrepresentation debate.

In the next section, I will discuss the basic tenets of the HOT theory and the kind of misrepresentation that is the subject of the debate. In section 3, I will present the phenomenon of change blindness and explain in detail why it is inadequate as evidence of

<sup>2</sup> Empirical data as well have been used to argue that the HOT theory is wrong (cf. Kozuch 2014).

misrepresentation. In section 4, I will consider three lines of reply that may be pursued in the attempt to make CB relevant to misrepresentation. I conclude that none of the three replies are successful and that CB is not evidence for misrepresentation.

## 2 HOT theory and misrepresentation

The issue of misrepresentation has been exhaustively explicated by Karen Neander (1998). Neander suggests a scenario in which three individuals (later dubbed ‘Neander’s Triplets’ by Ned Block in his 2011) have identical HOTS but differ in the lower-order states of those HOTS. All three have a HOT describing themselves as seeing something red. However, only the first individual has a target first-order state corresponding to seeing something red. The second individual has a target first-order state of seeing something blue, and the third individual has no target first-order state at all. In Neander’s terms the HOT of the second individual is *mildly* misrepresenting the first-order state, when it describes the individual as seeing something red rather than blue. In the third individual the HOT is a *radical* misrepresentation because there is no relevant first-order state at all. What Neander intends to show with the scenario is that according to the higher-order accounts of consciousness, the three individuals have identical conscious experiences. Seen from the first-person perspective it does not matter whether the relevant first-order state is there, is mildly misrepresented, or is not there at all. The three possibilities are subjectively indistinguishable.

The supposed theoretical implications of the possibility of misrepresentation, have incurred an extensive and lengthy debate. As mentioned above, the defendants of the HOT theory have begun supporting their argument with empirical data (Lau and Brown *to appear*, Lau and Rosenthal 2011a, Rosenthal 2012). One such piece of empirical data is change blindness.

## 3 Change blindness

The most well-known examples of CB are the so-called flicker paradigms pioneered by Rensink, O’Regan and Clark (Rensink, O’Regan and Clark 1997). In addition to these there is a wide variety of other

approaches in the field (see e.g. Beck, Rees, Frith and Lavie 2001, Busch 2010, Busch, Fründ and Herrmann 2010, Simons and Ambinder 2005, Simons and Rensink 2005), but I will limit myself to the flicker paradigms in the following.

In the flicker paradigm the subject is presented with two pictures separated temporally by a distractor screen. The two pictures differ in some respect, and the task of the subject is to detect what is changing. The change may concern a quite salient element of the two pictures and still have a strong experimental effect, e.g., a change of color, or the complete disappearance of a prominently featured object. It is hypothesized that the cause of the CB effect in the flicker paradigms is the distractor screen disrupting the pre-change representation. The distractor screen supposedly disrupts the visual trace of the pre-change picture in a way that precludes comparison to the post-change picture. Call this the *no comparison* (NC) interpretation of CB effect.<sup>3</sup> Now, it seems that because of the disruption, the flicker paradigms are ill-suited as evidence for misrepresentation. The reason they are ill-suited is that if the disruption between the pre-change and post-change representations fully explains the effect, then the flicker paradigms do not license any conclusions about misrepresentation. The reason why they do not license such conclusions is that according to the NC interpretation, the subject correctly represents the post-change first-order visual state, but is unable to compare it to his previous (pre-change) visual state. Plainly, on the NC interpretation there are no discrepancies between first-order states and HOTs, hence, no misrepresentation. Remember that misrepresentation occurs if there is a discrepancy between the HOT and the mental state it is about. Because, on the flicker paradigm the assumption is that visual representations are correct and the failure of the subjects is taken to consist in a failure to compare the stimuli, the change blindness paradigm that the defendants of the HOT theory prefer to cite in support of the possibility of misrepresentation is *saccade-induced change blindness* (SICB). Usually, the research of John Grimes on SICB (e.g. Grimes 1996) is brought forth, and this will be the focus of my assessment.

<sup>3</sup> The NC interpretation has received philosophical treatment by e.g. Dretske (2004) and Tye (2010).

In SICB the experimenter takes advantage of the fact that during saccades the subject is effectively blind. Using eye-tracking to determine the onset of saccades it is possible to time the switch in visual stimulus to occur during saccades. By changing the stimulus during saccades the SICB can replicate the experimental effects of other CB paradigms. Importantly, the SICB paradigm is supposed to avoid the NC interpretation because in this paradigm there is no distractor screen. If there is no distractor screen, then there cannot be any disruption of the pre-change representation preventing comparison of the pre- and post-change stimuli. It sometimes is speculated that smearing of the visual signal caused by rapid eye movements or brief cutoff of visual input during saccades have the same disruptive effect as a distractor screen (Simons and Ambinder 2005). Of course, if this is the case, the data from SICB will be subject to the NC interpretation and would be ill suited as evidence of misrepresentation. Since none of the arguments in this paper depend on this, we can concede the proponents of misrepresentation that saccades do not disrupt the pre-change state.

For misrepresentation to arise there must be a discrepancy between the first-order state and the HOT that represents it. In the SICB paradigm the content of the HOT is fixed by the state that the individual is aware of herself as being in. This means that the content is what the individual is conscious of, and the relevant first-order state is what the individual actually experiences herself as seeing. Rosenthal brings out the discrepancy that would cause misrepresentation by describing the situation as follows: “Because retinal input to visual cortex resumes after saccades, first-order states in visual cortex presumably did change in ways that reflected the change in display, despite participants’ reporting no awareness of such change.” (Rosenthal 2012: 8). Here Rosenthal claims that there is reason to think the post-change state is represented in the visual system. I agree with Rosenthal that there is good reason to believe a new first-order state occurs in the early visual system after the saccade. The conclusion Rosenthal draws is that since the subject does not report experiencing any change of states, he or she misrepresents the post-change first-order visual state. Thus, he claims that there is a discrepancy between the occurring HOT and the first-order visual state caused by the post-change stimulus. Because subjects do not detect changes

in the SICB paradigm, the state that they are aware of themselves as being in is the pre-change state. Evidently, the pre-change state does not correspond to the post-change state they actually are in.

However, why is it not the case that the subjects are in the states they report themselves as being in? Rather than inferring misrepresentation in the case of SICB, one might instead conclude that the HOT represents the pre-change state, and does so correctly. This conclusion is possible once we allow that the subject may have both the pre-change and the post-change states. There is no independent reason to think that the pre-change state has disappeared completely. Importantly, the pre-change state must exist in some form, although possibly not in the early visual system. If the pre-change state did not still exist in the system, the subject would not be able to perform the comparison of pre- and post-change states and consequently succeed in the change blindness task. We know that subjects are able to perform the change blindness task successfully. That they can do so indicates that the pre-change state is present in some form (see e.g. Simons and Rensink 2005 for some considerations on this).

This alternative interpretation can explain the data without positing the occurrence of misrepresentation. The explanation seems reasonable from an evolutionary point of view. The external world does not usually change on the miniature timescale on which saccades occur. It is unlikely that the visual system faced evolutionary pressure to generate a new HOT after every saccade. One might object that this consideration is tendentious because the opposite claim appears equally probable. The objection would be that it is equally probable that the visual system faced great selective pressure to stay up to date, and thus the “refresh rate” for HOTs would be very high. Yet, while it certainly is likely that there has been selective pressure on the refresh rate for HOTs, this claim is not a viable counter argument in the present case. The reason is that on the miniature timescale the generation of new HOTs is driven by visual transients (i.e. fast changes in luminance or color in the retinal image produced by sudden appearance, disappearance, or motion of objects). Most change blindness paradigms is set up to mask these transients in different ways. Since we know that retinal input resumes after saccades (cf. the quote from Rosenthal 2012 above) it appears that if the refresh rate of HOTs was as high as the objection has it, then change

blindness phenomena would not be possible.

On the alternative interpretation I propose, the subject does not have a HOT about the post-change state and consequently is in the same conscious state he or she was in before the saccade. One might protest that this interpretation entails that the HOT represents the pre-change state, even though this state no longer exists, viz., that the HOT survived the saccade but the pre-change first-order state did not. However, this is not the claim I want to make.

I suggest that the pre-change state may survive the saccade, and consequently that the HOT (whether a new one or one that was generated before the saccade) that is about this state can be a correct representation of it.<sup>4</sup> This is in line with the consideration that both pre- and post-change states can exist in the system. Now, because retinal input to visual cortex resumes after saccades, one might conclude that the pre-change state has been overwritten, and then my claim that it may survive would be false. However, to emphasize, I am not claiming that the pre-change state necessarily lingers as raw sensory information in the early visual areas. It is possible that any information in these areas is overwritten once retinal input resumes after saccades, as Rosenthal suggests. Nevertheless, there is mounting evidence that visual information about previously attended scenes is retained in other areas (see e.g. Hollingworth and Henderson 2002). In fact, it is reasonable to think that such information is retained in memory and plays different roles, say, for learning from perception and action as well as the recall of visual scenes in the environment. It might also linger in short term memory in the absence of salient or competing external input.

Moreover, from a theoretical standpoint, defendants of the HOT theory cannot pursue this line of thought because it allows for the

<sup>4</sup> One could object here that the proponents of HOT theory need not be committed to a new HOT being generated after the saccade. Certainly, some versions of HOT theory (e.g. Van Gulick 2004) could pursue this reply. However, on most versions of HOT theory (including the one proposed by Rosenthal) a HOT is individuated at least partly by its content. This means that a HOT cannot change its content, so claiming that the pre-change HOT survives and changes its content is not an available position. Furthermore, since it introduces no discrepancies between the HOT and its content, it is unclear how this position would support misrepresentation.

NC interpretation. The SICB paradigm was initially preferred over the flicker paradigms, because it could avoid the NC interpretation. On the NC interpretation, there is no discrepancy between the first-order state and the HOT and thus nothing follows with respect to misrepresentation. Thus, from a theoretical standpoint, claiming that the pre-change state has disappeared does not salvage the SICB data as evidence of misrepresentation.

To summarize, I agree with Rosenthal that because retinal input resumes after the saccade, there is reason to believe the post-change state exists. My claim is that the pre-change state exists in addition to the post-change state. Rosenthal, along with most participants in the misrepresentation debate, believe that an individual can have unconscious mental states (see e.g. Rosenthal 1997: 731-734, 2000a: 203, Rosenthal 2004a: 162). This means that on their view, we can allow that after the saccade both the pre- and post-change states exist simultaneously as first-order states, perhaps in different cortical areas.

The view that the pre-change state exists does not as such seem very controversial. In fact, it appears that defendants of the possibility of misrepresentation have to endorse this view in order to avoid the NC interpretation. Therefore, the interpretation that a HOT represents the pre-change state is a viable position.

#### 4 Three possible replies

It is hard to see how the defendants of the possibility of misrepresentation might counter the alternative interpretation of the SICB data. To do this, it seems they need to provide evidence of the representational relations that actually obtain when misrepresentation occurs. Remember that misrepresentation means that the HOT either represents another first-order state than the subject experiences (mild misrepresentation) or it does not represent any first-order state at all (radical misrepresentation). Of course, in the cases where subjects do detect the changes, we can know that they are correctly representing the post-change state, and have succeeded in comparing it to the pre-change state. However, there appears to be no way of determining what representational relationship obtains in the cases where changes are not detected. Furthermore, if evidence of the

representational relations nevertheless could be produced, it would be sufficient to determine if misrepresentation was possible, and the SICB data would be superfluous.

Since we have no evidence about the representational relation in case of SICB, there is no way to know whether the HOT represents the pre-change or the post-change state. Call this the *interpretation problem*. Next, I will consider three possible lines of reply to the interpretation problem that the proponents of misrepresentation might pursue. The replies attempt to save SICB as evidence of misrepresentation in three different ways. The first line of reply questions whether it is reasonable to say that the HOT is about the pre-change state. The second line of reply accepts that there is a problem but attempts to show that, even if the HOT is representing the pre-change state, misrepresentation still occurs. The third line of reply objects that due to a proviso in the HOT theories, the HOT cannot be allowed to represent the pre-change state in SICB, as I have suggested.

#### 4.1 The first line of reply

The first line of reply denies that the HOT is about the pre-change state. If this is true, we can infer that the HOT is about the post-change state, and misrepresentation follows. The reply can be developed in two ways; I will argue that neither is successful. The first way to develop it is to claim that the HOT in fact *cannot* be about the pre-change state. Conceivably the motivation for this claim would be the belief that the pre-change state somehow is degraded or ceases to exist. Do we have any reason to believe that the pre-change state disappears? There does not seem to be any empirical evidence in support of this. On the contrary, if the evidence (cf. Hollingworth and Henderson 2002) pertaining to the storage of visual information outside of the early visual areas is correct, this is reason to believe the pre-state exists. However, a more important consideration is the following: If SICB is to count as evidence of misrepresentation, there *must not* be any reason to believe that the pre-change state has disappeared. Remember that SICB was favored because the traditional flicker paradigm was subject to the NC interpretation. The NC interpretation explained the experimental effect by positing that the pre-change state was disrupted. Because of the disruption, the

subjects could not perform the necessary comparison of the visual stimuli to detect the changes.

As against the explanation of change blindness by failed comparison, it might be objected that change detection does not require comparison. The argument would be that change monitoring is sufficient. Suppose a subject is monitoring a green screen in order to detect when it changes its colors. It would seem that the subject merely needs to represent the current color and register that it disappears to report that the color has changed. Comparison would not be necessary. Similarly, once the color has changed, it would seem that the subject does not have to represent the pre-change color of the screen to report that there was a change. However, the objection misses the point. In the traditional change blindness tasks, the subjects are not informed about what aspects of the visual stimuli will change. This means that they cannot key their attention to a single feature and wait for it to change. Subjects are required to report what has changed, and because they are unaware of which object or property will change, they have to monitor as many features of the display as possible and actively search for a change. In addition, to succeed on the task, it is not sufficient to assert that ‘something’ has changed or that there was a change. Subjects sometimes report a “feeling of change” without being able to pinpoint what has changed and in which respect (see e.g. Simons and Rensink 2005). This behavior is taken to indicate the presence of implicit change detection. To actually know that a change has taken place requires being able to describe what the change consisted in<sup>5</sup>.

The other way to develop the first line of reply is to argue that the post-change state constitutes the *relevant* first-order state. Supposedly, it is the relevant state, because it is the state that actually corresponds to the concurrent visual input. On this approach, the argument is that in SICB, misrepresentation occurs because the HOT fails to represent the *relevant* first order state. As I see it, there are two problems with this view. The first problem is that it introduces a criterion of relevance into a theory that aims to explain the difference between conscious and unconscious mental states. The criterion does not receive any independent support, and since it serves

<sup>5</sup> I am thankful to an anonymous referee of this journal for raising this issue.

no other purpose in the theory than to secure SICB as evidence of misrepresentation, it is *ad hoc*. Furthermore, introducing the notion of relevance in this manner consists in a significant modification of the HOT theory. A seemingly normative or at least arbitrary criterion would be an uncomfortable amendment to a theory that seeks to provide an objective explanation of consciousness.

The second problem for this way of developing the first line of reply is that introducing relevance appears to change the notion of misrepresentation. It amounts to a bait and switch maneuver in order to save the SICB data as evidence of misrepresentation. The notion of misrepresentation originally was defined in terms of the discrepancy between a HOT and its target state. Suggesting that misrepresentation occurs when a HOT is not representing the relevant state appears to introduce a different notion. What exactly such a different notion of misrepresentation might consist in is moot, and a separate question. What matters here is that this is not the notion of misrepresentation that SICB was introduced to support.

#### 4.2 The second line of reply

The second line of reply to the interpretation problem is to argue that the conclusion still follows. This line concedes that there is no way of knowing whether the HOT represents the pre-change state correctly or not. However, SICB nevertheless presents evidence of misrepresentation. We may now ask: In what way the HOT misrepresents its target in such a case? The answer is that the HOT misrepresents the pre-change state in the sense that it represents its content (the pre-change visual stimulus) as *externally present*, when the visual stimulus in fact has been replaced by the post-change stimulus. Supposedly, misrepresentation obtains because the HOT describes the individual as seeing *x*, when the individual in fact is seeing *y*.

It may appear as if the second line of reply preserves the notion of misrepresentation as a discrepancy occurring between a HOT and the state it is about. But this appearance is misleading. First, the question is in virtue of *what* the previously correctly representing HOT suddenly became a misrepresenting HOT? The HOT did not change its contents, and neither did the state that it was about. The change that turned the correctly representing HOT into a misrepresentation

was, it seems, external to the subject. The change was to the visual stimuli, not to the visual experience. This means that the misrepresentation occurs because there is a change in how the mental content corresponds to the external world. Plainly, the alleged misrepresentation occurs because the state that the individual is conscious of herself as being in does not correspond to external reality. Like on the first line of reply, this move appears to change the notion of misrepresentation. On the original definition, the discrepancy concerns the relation between the subject's experience and her first-order visual state.

In addition, notice that the understanding of misrepresentation as a discrepancy between (conscious) mental content and the external world is uncontroversial. The view that we sometimes may be wrong about states of the external world is uncontroversial. I may occasionally see a cow and think it is a horse, but this is inconsequential to the notion of misrepresentation that SICB is introduced to support. Thus, it appears that the second line of reply fares no better than the first line of reply. Both lines of reply save the argument from SICB by changing the notion of misrepresentation. While these replies may secure that SICB remains informative about the nature of misrepresentation, they do so at the cost of changing the notion of misrepresentation that the SICB data originally was intended to be evidence of. Crucially, these changes apparently render the notion of misrepresentation defended irrelevant to the debate in which SICB was supposed to figure as an argument.

### 4.3 The third line of reply

The third line of reply argues that a proviso of the HOT theories prevents the interpretation of the SICB data that has been presented here. In original formulations of the HOT theory a mental state is conscious when it is *accompanied* by a HOT (e.g. Rosenthal 1997: 741). This claim is usually taken as the condition that a HOT must be roughly simultaneous with the mental state it is about. Call this the *simultaneity criterion*. The simultaneity criterion is well justified. The basic motivation is that it is undesirable that a HOT makes an individual conscious of herself as e.g. 'seeing her grandmother now', when she really was seeing her grandmother yesterday. Our con-

scious awareness should customarily be about whatever our first-order states currently represent.

There are two ways of responding to the objection based on simultaneity. The first response consists in challenging the simultaneity criterion. The second response consists in showing that even if the HOT is technically representing a state of the past it nevertheless satisfies the simultaneity criterion. I will consider these responses in turn.

The first response challenges the scope of the simultaneity criterion in the following way. It seems that there is a tension between on the one hand, holding the simultaneity criterion and on the other, embracing the possibility of radical misrepresentation. If it is allowed that HOTs can occur in the absence of first-order states, how is it possible to hold that a HOT must be simultaneous with the first-order state it is about? Simultaneity is a two-place relation, and therefore depends on the existence of two relata. This problem is akin to another issue in the debate on misrepresentation. It has been argued (e.g. Block 2011a) that if the HOT theories view conscious states as the upshot of a two-place representational relation between a HOT and a first-order state, then conscious states cannot occur in radical misrepresentation because one of the relata is missing. With respect to this issue, the proponents of HOT theory have replied that conscious states do not depend on the existence of such a two-place relation. A HOT is in itself sufficient to bring about a conscious state (Rosenthal 2011). This reply works because a representational relation may take inexistent objects as one of the relata, as when one thinks of the Easter Bunny. However, this kind of reply cannot help solve the simultaneity problem, because a thing cannot be simultaneous with something that does not exist. The simultaneity criterion must be framed as a relation between an existing first-order state and a HOT.

Given that the simultaneity criterion has a legitimate role to play in the HOT theory, how might it be saved? A straightforward possibility comes to mind: The simultaneity criterion is only intended to hold in the *normal case*, which means that the possibility of misrepresentation, falling outside of the normal case, is in the clear. However, this view precludes the third line of reply, because that line of reply presupposes that the simultaneity criterion holds for the SICB data.

Arguably, cases of SICB do not count as normal. Thus, if the scope of the simultaneity criterion is narrowed to the normal case, apparently, the criterion cannot be used to argue against the interpretation of CB put forward in this paper, viz., that the HOT is correctly representing the pre-change state.

The second response to the reply in terms of simultaneity emphasizes that the assumption that the HOT represents a pre-change state does not entail that it cannot be simultaneous with the pre-change state. Empirical research on conscious visual awareness suggests roughly 100 milliseconds (ms) of processing is required, after a visual stimulus is presented to the retina, before the subject can become consciously aware of it (e.g. Lamme 2003: 15). Since the presence of a HOT reveals itself by the occurrence of conscious visual awareness, this means that it takes at least 100 ms for a HOT to be generated. Thus, it seems that a latency of 100 ms between visual input and the occurrence of a HOT is the minimum amount of time that satisfies the simultaneity criterion. Hence, the lower threshold of simultaneity roughly would be 100 ms, because below this threshold no HOT is generated.

The simultaneity criterion was introduced as a proviso against the existence of HOTs about first-order states that belong to the past. Supposedly, according to the objection from simultaneity, the pre-change states of the CB paradigm are examples of states that belong to the past, because too long a time has passed since they emerged — “too long” in comparison to first-order states that have emerged later such as the post-change states. The moment that delineates when an episode in time starts to count as “too long”, can be conceived of as the upper threshold for the simultaneity criterion. This means that for simultaneity to hold, the HOT also must occur below certain latency. The exact latency of this upper threshold is an open question, but it is reasonable to assume that there is some window to allow for individual differences in neural processing. However, any estimate of the upper threshold is likely to be arbitrary at this point. The window might span very little time, or as much as several hundred milliseconds. The exact range is inconsequential for the argument here. The important point is that there is some wiggle room in exact latency within which simultaneity can be said to obtain. Simultaneity thus does not seem entail that the HOT and the first-order state

emerge at the same point in time or that they overlap in time.

Now, consider that the timescale on which saccades occur is roughly between 5 and 80 ms (Grimes 1996: 90). Consequently, a HOT existing after a saccade still will be below the lower threshold of the simultaneity criterion because it will be maximally 80 ms removed from the visual input it is about. Furthermore, on the reasonable assumption that there is a window above the lower threshold, within which simultaneity holds, it becomes even more plausible that a HOT could represent the pre-change state while still satisfying the simultaneity criterion.

To conclude, I have examined three lines of reply against my suggested interpretation of the SICB data. The upshot of this examination is that none of these lines of reply can sustain SICB as evidence of misrepresentation. The first two lines of reply fail to preserve the notion of misrepresentation that is at stake in the debate. The third line is rejected on the basis of two independent arguments. Thus, the interpretation problem remains: Since we do not have any reliable evidence about the representational relation in the case of saccade induced change blindness, there is no way of knowing whether the HOT represents the pre-change or the post-change state. Pending further argument or evidence, it seems that whether saccade induced change blindness can be counted as evidence of misrepresentation remains an open question.<sup>6</sup>

Asger Kirkeby-Hinrup  
Dept. of Philosophy and Cognitive Science  
Lund University  
Box 192, 22100 Lund, Sweden  
asger.kirkeby-hinrup@fil.lu.se

### *References*

- Balog, Kati. 2000. Comments on David Rosenthal's "Consciousness, content, and metacognitive judgments." *Conscious and Cognition* 9(2 Pt. 1): 215-219; discussion: 231-242.
- Beck, Diane M., Rees, Geraint, Frith, Christopher D., and Lavie, Nilli.

<sup>6</sup> I am indebted to Prof. Ingar Brinck for guidance and insightful discussion and Prof. David Rosenthal for inspiring correspondence in developing this paper.

2001. Neural correlates of change detection and change blindness. *Nature Neuroscience* 4(6): 645-650.
- Berger, Jacob. 2014. Consciousness is not a property of states: a reply to Wilberg. *Philosophical Psychology* 27(6): 829-842.
- Block, Ned. 2011a. The higher order approach to consciousness is defunct. *Analysis* 71(3): 419-431.
- Block, Ned. 2011b. Response to Rosenthal and Weisberg. *Analysis* 71(3): 443-448.
- Busch, Niko. 2010. Objective and subjective measures of change blindness: where are the real pitfalls? *Journal of Cognitive Neuroscience* 22(9):1903-1905.
- Busch, Niko, Fründ, Ingo, and Herrmann, Christoph. 2010. Electrophysiological evidence for different types of change detection and change blindness. *Journal of Cognitive Neuroscience* 22(8): 1852-1869.
- Carruthers, Peter. 1998. Natural theories of consciousness. *European Journal of Philosophy* 6(2): 203-222.
- Dretske, Fred. 2004. Change blindness. *Philosophical Studies* 120(1-3): 1-18.
- Gennaro, Rocco J. 1996. *Consciousness and Self-consciousness: A defense of the higher-order thought theory of consciousness*. Amsterdam: John Benjamins Publishers.
- Gennaro, Rocco J. 2012. *The consciousness paradox: Consciousness, concepts, and higher-order thoughts*. Cambridge, MA: MIT Press.
- Grimes, John A. 1996. On the failure to detect changes in scenes across saccades. In *Perception*, ed. by Kathleen Akins: 89-110. New York: Oxford University Press.
- Hollingworth, Andrew, and Henderson, John M. 2002. Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance* 28(1): 113.
- Kirkeby-Hinrup, Asger. 2014. Why the rare Charles Bonnet cases are not evidence of misrepresentation. *Journal of Philosophical Research* 39: 301-308.
- Kozuch, Benjamin. 2014. Prefrontal lesion evidence against higher-order theories of consciousness. *Philosophical Studies* 167(3): 721-746.
- Lamme, Victor A. F. 2003. Why visual attention and awareness are different. *Trends in Cognitive Sciences* 7(1): 12-18.
- Lau, Hakwan. 2007. A higher order Bayesian decision theory of consciousness. *Progress in brain research* 168: 35-48.
- Lau, Hakwan and Brown, Richard. (To Appear). The emperor's new phenomenology? The empirical case for conscious experiences without first-order representations. In *Themes from Block*, ed. by Adam Pautz and Daniel Stoljar. Cambridge, MA: MIT Press.
- Lau, Hakwan, and Rosenthal, David M. 2011a. Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences* 15(8): 365-373.
- Lau, Hakwan, and Rosenthal, David M. 2011b. The higher-order view does not require consciously self-directed introspection: response to Malach. *Trends in Cognitive Sciences* 15(11): 508-509.
- Matey, Jennifer. 2006. Two HOTs to handle: the concept of state consciousness in the higher-order thought theory of consciousness. *Philosophical Psychology*

- 19(2): 151-175.
- Matey, Jennifer. 2011. Reduction and the determination of phenomenal character. *Philosophical Psychology* 24(3): 291-316.
- Neander, Karen. 1998. The division of phenomenal labor: a problem for representational theories of consciousness. *Nous* 32(S12): 411-434.
- Rensink, Ronald A., O'Regan, J. Kevin, and Clark, James A. 1997. To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science* 8(5): 368-373.
- Rosenthal, David M. 1997. A theory of consciousness. In *The Nature of Consciousness: Philosophical Debates*, ed. by Ned Block, Owen Flanagan and Güven Güzeldere: 729-753. Cambridge, MA: MIT Press.
- Rosenthal, David M.. 2000a. Consciousness, content, and metacognitive judgments. *Consciousness and Cognition* 9(2): 203-214.
- Rosenthal, David M. 2000b. Metacognition and higher-order thoughts. *Consciousness and Cognition* 9(2): 231-242.
- Rosenthal, David M. 2002a. Explaining consciousness. In *Philosophy of Mind: Classical and contemporary readings*, ed. by David Chalmers: 406-421. Oxford: Oxford University Press.
- Rosenthal, David M. 2002b. How many kinds of consciousness? *Consciousness and Cognition* 11(4): 653-665.
- Rosenthal, David M. 2004a. Being conscious of ourselves. *Monist* 87(2): 159-181.
- Rosenthal, David M. 2004b. Varieties of higher-order theory. In *Higher-order Theories of Consciousness*, ed. by Rocco J. Gennaro: 19-44. Amsterdam: John Benjamins Publishers.
- Rosenthal, David M. 2011. Exaggerated reports: reply to Block. *Analysis* 71(3): 431-437.
- Rosenthal, David M. 2012. Higher-order awareness, misrepresentation and function. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1594): 1424-1438.
- Sebastián, Miguel Ángel. 2014. Not a HOT dream. In *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, ed. by Richard Brown: Vol. 6, 415-432. New York: Springer Press.
- Shepherd, Joshua. 2013. Why Block can't stand the HOT. *Journal of Consciousness Studies* 20(3-4): 183-195.
- Simons, Daniel J., and Ambinder, Michael S. 2005. Change blindness: theory and consequences. *Current Directions in Psychological Science* 14(1): 44-48.
- Simons, Daniel J., and Rensink, R. A. 2005. Change blindness: past, present, and future. *Trends in Cognitive Sciences* 9(1): 16-20.
- Tye, Michael. 2010. Attention, seeing, and change blindness. *Nous* 20(1): 410-437.
- Van Gulick, Robert. 2004. Higher-order global states (HOGS): an alternative higher-order model. In *Higher-order Theories of Consciousness*, ed. by Rocco J. Gennaro: 67-93. Amsterdam: John Benjamins Publishers.
- Weisberg, Josh. 2008. Same old, same old: the same-order representation theory of consciousness and the division of phenomenal labor. *Synthese* 160(2): 161-181.

- Weisberg, Josh. 2010. Misrepresenting consciousness. *Philosophical Studies* 154(3): 409-433.
- Weisberg, Josh. 2011. Abusing the notion of what-it's-like-ness: a response to Block. *Analysis* 71(3): 438-443.
- Wilberg, Jonah. 2010. Consciousness and false HOTs. *Philosophical Psychology* 23(5): 617-638.