January 2013

The Resilience of a Refined Higher-Order Thought Theory of Consciousness

Lee-Anna T. Sangster
The University of Western Ontario

Supervisor
Robert Stainton
The University of Western Ontario

Graduate Program in Philosophy

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

THE RESILIENCE OF A REFINED HIGHER-ORDER THOUGHT THEORY OF CONSCIOUSNESS

(*Spine Title*: The Resilience of a Refined HOT Theory of Consciousness)
(*Thesis Format*: Integrated Article)

by
Lee-Anna T. Sangster
Graduate Program in Philosophy

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
Dr. Robert Stainton	Dr. Andrew Brook
Supervisory Committee	Dr. Bertram Gawronski
Dr. Andrew Botterell	Dr. John Nicholas
	Dr. Chris Viger

The thesis by

Lee-Anna Theresa Sangster

entitled:

The Resilience of a Refined Higher-Order Thought Theory of Consciousness

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date

Chair of the Thesis Examination Board

Abstract

This dissertation consists of three independent papers, each defending the Higher-Order Thought (HOT) Theory of Consciousness against a different objection. First the HOT theory is defended against the Theory of Mind (TOM) Objection. Since the HOT theory requires that a subject be able to represent mental states in thought in order to have mental states that are conscious, objectors argue from empirical evidence that few creatures pass TOM tests to the conclusion that few creatures must be capable of having conscious mental states according to the HOT theory. The counter-intuitiveness of this claim is then taken as reason for rejecting the HOT theory. I argue that this objection is based on a false assumption - that the requirements of successful TOM test performance parallel the requirements outlined by the HOT theory. Since this assumption is false, we can reject the objection. In the second paper, I defend the HOT theory against the Phenomenal Character Argument. Objectors argue that the HOT theory must be rejected because it incorrectly characterizes the phenomenal character affiliated with basic state consciousness as necessarily involving a consciousness *of the fact that one has a particular mental state*. I argue that the theory cannot provide this characterization of phenomenal character because the theory cannot say that someone becomes conscious of what her unconscious HOTs represent. Since the objection rests on an incorrect interpretation of the theory, we can reject the objection. In the final paper, I defend the HOT theory against the Misrepresentation Objection. Here objectors accuse the HOT theory of presenting necessary and sufficient conditions for conscious states that turn out to be incompatible in empty HOT cases (cases wherein one's HOT misrepresents the states one is instantiating). I argue that the conditions are actually each part of separate explanations of separate sorts of consciousness, that neither of the separate explanations has internally incompatible conditions, and hence that the objection is based on an equivocation on two senses of the phrase 'conscious state'. Since the objection is based on this error, we can reject the objection.

Keywords: Higher-Order Thought, Consciousness, David Rosenthal, Theory of Mind, Metacognition, Phenomenal Character, Empty HOTs, Metaphysics of Mind.

For Bill Heintz,
who taught me that life demands humour, wit, and grit

And for Chris Nicholl,
who is, and always will be, my 'Crazy Love'

Acknowledgements

Our work cannot help but be influenced by those around us; this project has been no exception. In particular I would like to thank Saad Anis, Guillaume Beaulac, Jacob Berger, Sheldon Chow, Matthew Ivanowich, Jonathan Life, Corey Mulvihill, Myrto Mylopoulos, James Overton, Lisa Pelot, Nicholas Ray, Chris Young, and the rest of my friends and colleagues at the University of Western Ontario for their helpful conversations and moral support. I would also like to thank those who offered valuable feedback, at various stages of this project's development, when drafts were presented at the WCPA, the CPA, the ASSC, the SPP, the PPRG, and the UWO PGSA Graduate Speaker Series.

I would like to acknowledge my gratitude for financial support provided by The Ontario Graduate Scholarship Program, The Richard A. Harshman Scholarship, The Social Sciences and Humanities Research Council (via a grant awarded to Professor Robert Stainton), and The Western Graduate Thesis Research Award Program.

I would also like to thank the 2006-2012 UWO Philosophy Department intramural volleyball teams for providing me with a much needed break from academic pursuits!

For their support and inspiration, if not on this project in particular then in my development as a scholar in general, I thank professors Robert Batterman, John Bell, Andrew Botterell, Andrew Brook, Lorne Falkenstein, Bertram Gawronski, Philip Hanson, William Harper, Thomas Lennon, Ausonio Marras, Paul Minda, Wayne Myrvold, John Nicholas, David Rosenthal, William Seager, and Chris Viger.

And finally, I am forever indebted to my supervisor, Robert Stainton. Thank you for your unwavering support, your unbelievable wisdom, your untiring patience, and your unexpected master classes on fine tequila and carp fishing. I have grown so much from knowing you; you have truly been a great mentor and a great friend. Thank you.

Table of Contents

List of Tables

Introduction

The Higher-Order Thought (HOT) Theory of Consciousness is a fascinating hypothesis about the underpinnings of mental state consciousness and the nature of conscious experience. As it is most commonly understood, the theory explains our conscious mental states in terms of a higher-order awareness we come to have of those states and it explains our phenomenal experiences in terms of the way this awareness characterizes our mental environment. Specifically, HOT theorists argue that a mental state's consciousness is constituted by one's representing oneself as being in a lower-order mental state by means of forming an appropriate higher-order thought.

As the title of this dissertation suggests, I intend to present and defend a refined version of the HOT Theory. I achieve this by exploring the theory through the lens of three different objections: The Theory of Mind Objection, the Phenomenal Character Argument, and The Misrepresentation Objection. As we will see, each objection is directed toward a different aspect of the HOT theory and, as I will argue, each objection can be defeated by bringing out some of the often unnoticed subtleties within the theory. Hence what results is a refined version of the HOT theory, in the sense that we replace our usual understanding of the theory with a more careful and nuanced interpretation.

Since the goal is to bring out the subtle nuances of the HOT theory and to demonstrate how these nuances help the theory address various objections, I make a few important methodological choices. First, because the details matter in such a project, and because different theorists each tend to promote a slightly different variation of the theory, within this dissertation I focus exclusively on one of the most accepted and well-

known versions of the HOT theory – David Rosenthal's (1986, 2005) Actualist HOT Theory.[1] Only by sacrificing breadth for depth can one explore the nuances that I argue are important.

Second, I choose to structure my discussions from within the conceptual framework with which the HOT theory is *most commonly* understood and to express any refinements that emerge with new terminology. This means, for example, that though Rosenthal has suggested his own refinements to the theory when responding to some of these objections, I will present an independent account of how the theory ought to be interpreted.[2]

I choose this option because it is important to me that the objectors be able to follow my path to a new, refined interpretation of the theory and I feel the best way to achieve this is to start from common ground and show how the new refinements develop out of the conceptual framework with which the objectors are most familiar.

With these notes in mind, let me present an overview of the three papers that constitute this dissertation.

In the first paper I defend the HOT theory against the Theory of Mind Objection.[3] Since HOT theorists argue that one must represent one's lower-order mental state with an appropriate HOT, in order for that lower-order state to be conscious, their account entails that for a person to have any conscious mental states at all, she must have the general capacity to represent her mental states in thought. Many see this requirement as a reason

---

[1] For examples of other Higher-Order accounts of consciousness see Gennaro (1996), Carruthers (2000) and Lycan (1996).

[2] For example, though Rosenthal often describes the property of mental state consciousness as being a *relational* property of mental states (see Rosenthal (2002) for example), we learn from his responses to the Misrepresentation Objection that this characterization of state consciousness must be refined (see Rosenthal (2003, 2011) for example).

[3] For examples of this objection in the literature see Dretske (1995), Ch. 4 or Seager (2004).

to reject the HOT theory. Specifically, proponents of the Theory of Mind objection argue that very few creatures on earth seem capable of representing mental states in thought, and hence that the HOT theory entails that very few creatures on earth are capable of having conscious mental states at all. The counter-intuitiveness of this claim is taken as reason for rejecting the HOT theory as an explanation of mental state consciousness.

To support the claim that few creatures have the capacity to represent mental states in thought, objectors appeal to evidence from Theory of Mind (TOM) research conducted in Developmental and Comparative Psychology, hence this objection is referred to as the *Theory of Mind Objection*. A person is said to have a *theory of mind* when she is able to successfully predict or explain someone else's behaviour by attributing to that other person certain mental states.[4] Since studies reveal that most non-human animals and most human infants lack a competence in Theory of Mind, objectors reason from this evidence to the conclusion that those who fail these tests must lack the capacities the HOT theory requires for mental state consciousness.[5]

I argue that in order to take the failure of subjects on TOM tests as evidence against the HOT theory, however, one must assume that the requirements of successful performance on TOM tests parallel the very requirements for mental state consciousness outlined by the HOT theory. I call this the Parallel Requirements Assumption and I argue that it turns out to be false when assessed in relation to the standard verbal TOM tests that are taken as support for the TOM objection. In light of this fact I conclude that the TOM

---

[4] Premack and Woodruff (1978) originally introduced the term 'theory of mind'.

[5] For early studies demonstrating that human infants do not show evidence of these abilities until around the age of 4 years see Wimmer and Perner (1983) and Perner, Leekam, and Wimmer, (1987). For discussion of these abilities in nonhuman animals, consider the research on chimpanzees, a population considered among the most likely nonhuman population to have these abilities. Though researchers continue to debate about the theory of mind skills chimpanzees may or may not have, there has yet to be any uncontroversial evidence of false belief understanding in this population. For representative studies from both sides of the debate see Povinelli, Nelson, and Boysen (1990) and Hare, Call, and Tomasello (2001).

objection must be rejected. (I also show that we can reject a similar argument that might be created, this time in support of the HOT theory, on the basis of newly emerging evidence that young human infants can pass non-verbal TOM tests.[6] Since such an argument would also require adopting the Parallel Requirements Assumption, and since I argue that this assumption is not justified relative to these tests either, I conclude that this argument also must be rejected.)

In order to demonstrate that the Parallel Requirements Assumption is unjustified, I provide an interpretation of a rarely acknowledged aspect of the theory, namely, its account of phenomenal character. In particular, I argue that the HOT theory dictates that when a mental state, *M*, is represented by an appropriate HOT, and hence when *M* is state conscious, the creature instantiating this state will become conscious of what *M* represents. This allows me to show that the standard verbal TOM tests require cognition that is structurally more complex than what the HOT theory requires for mental state consciousness (and it allows me to show that new nonverbal TOM tests might require cognition that is structurally less complex than what the HOT theory requires for mental state consciousness). As it turns out then, TOM tests do not measure what the HOT theory requires for mental state consciousness, and hence TOM tests cannot support arguments about the plausibility of these requirements. The TOM objection must therefore be rejected.

The second paper is a defense of the HOT theory against Robert Lurz's (2003, 2006) Phenomenal Character Argument. It is also an opportunity to initiate a much needed discussion of the HOT theory's account of phenomenal character. According to

[6] For example, see Onishi and Baillargeon (2005) and Southgate, Senju, and Csibra (2007).

Lurz, the HOT theory accounts for phenomenal character by claiming that one necessarily becomes conscious *of the fact that one has that particular mental state* whenever one comes to have a mental state that is conscious. Lurz demonstrates, however, that this characterization of phenomenal character is false and so he concludes that we ought to reject the HOT theory as an explanation of phenomenal character.

I argue that Lurz's interpretation of the HOT theory must be incorrect, since the HOT theory *cannot* say that the phenomenal character affiliated with mental state consciousness involves one's becoming conscious of what one's (unconscious) HOTs represent. In light of this error I conclude that Lurz's objection must be rejected.

I end this paper by discussing an implication of my argument, namely, that there must be an important distinction between the account of *what grounds* mental state consciousness that is offered by the HOT theory and the account of *what it's like to have* conscious mental states that is offered by the HOT theory. These accounts are distinct, I argue, because only one can be provided in terms of higher-order representation. The idea that there is separation between these two accounts, and the details emerging about the HOT theory's account of phenomenal character, are two of the primary revisions that comprise the refined understanding of the HOT theory that I hope to deliver throughout the dissertation.

In the third paper I defend the HOT theory against Ned Block's (2011a, 2011b) version of the Misrepresentation Objection. The HOT theory claims that a mental state is conscious only if it is represented by an appropriate HOT, but the theory allows that a HOT might *mis*represent a person as being in a state that she is not in fact instantiating. Furthermore, the theory allows that, despite this misrepresentation, it would still seem to

the person that she is in the state she, in fact, is not instantiating.[7] Such cases of misrepresentation are known in the literature as empty HOT cases.

Block argues that these cases prove fatal to the HOT theory, because they reveal that the conditions the theory sets out as being necessary and sufficient for conscious states actually lead to incompatible predictions about the presence of conscious states in empty HOT cases. In particular, it would appear that the HOT theory is committed to there being something it's like for a creature, and hence committed to there being a conscious state in empty HOT cases, despite the fact that there is no instantiated lower-order mental state to which we might attribute the property of mental state consciousness. In light of this result, Block concludes that we cannot accept the explanation of conscious states provided by the HOT theory and hence that the HOT theory must be rejected.

I argue that Block's Misrepresentation Objection can be seen as being based on an equivocation of two senses of the phrase 'conscious state' – a state consciousness sense of this phrase and a new, subject consciousness sense of the phrase. I demonstrate how the HOT theory can be seen as providing separate explanations of both sorts of consciousness identified by these senses of 'conscious state' and I argue that neither of these explanations lead to incompatible predictions, as Block originally objected. In light of this fact I conclude that Block's Misrepresentation Objection must be rejected.

Here we see a result similar to the conclusion of the previous paper. Once again it appears that the HOT theory is providing distinct explanations of what constitutes state consciousness and of phenomenal character. In this paper I also provide a brief discussion of a potentially surprising consequence of this way of interpreting the HOT theory, namely, that the property we traditionally identify as mental state consciousness

[7] For example, see Rosenthal (2004), especially pg. 32-35 and Rosenthal (2005), especially pg. 217-218.

turns out actually to play a less important role in the theory's explanation of phenomenal character than might have been expected.

And thus we complete our outline of the arguments that comprise this dissertation. With this overview in hand, I would like to highlight some aspects of the refined HOT theory that I think emerge from these discussions. In order to bring light to this new way of interpreting the HOT theory, I will briefly review how each objection attacks a slightly different aspect of the theory and how each reply brings out a slightly different refinement to our understanding of the HOT theory.

First, we see that the TOM objection attacks the theory's metarepresentational requirement. My reply is not to deny that the theory has such a requirement, rather it is to demonstrate that, with a proper understanding of the HOT theory's account of phenomenal character, it becomes clear that the TOM tests cannot be measuring the metarepresentational capacities the HOT theory requires. Since the TOM tests do not track these abilities, however, a subject's performance on these tests cannot inform us about whether that creature has what it takes to have mental states that are conscious, according to the HOT theory. Hence the TOM Objection, which is based on the evidence from these TOM tests, can be rejected. What we learn from this paper, then, in terms of the refined HOT theory, is that there is a rich account of phenomenal character within the HOT theory that may have unnoticed explanatory power.

Second, we see that the Phenomenal Character Argument attacks what Lurz (2003, 2006) takes to be the HOT theory's account of phenomenal character. My reply is to argue that Lurz misrepresents the HOT theory's account of phenomenal character. By showing that the HOT theorists cannot provide the account Lurz takes them to offer, we

are able to reject the Phenomenal Character Argument. What we learn from this paper, then, about the refined HOT theory, is that the HOT theory cannot provide an account of phenomenal character in terms of higher-order representation, so the theory's account of phenomenal character must be importantly distinct from its account of what constitutes mental state consciousness.

Finally, we see that the Misrepresentation Objection attacks the HOT theory by questioning the role of mental state consciousness in the theory's explanation of phenomenal character. My reply is to demonstrate how, on our refined understanding of the HOT theory, mental state consciousness does turn out to be quite independent from phenomenal character. I also demonstrate, however, that there is an alternative account of phenomenal character provided by the theory, so I show that nothing is lost by separating these two aspects of the theory. What we learn from this paper, then, about the refined HOT theory, is that the HOT theory's account of mental state consciousness plays a less central role in the HOT theory's account of phenomenal character than was previously expected.

With this summary it becomes clear that a common thread is emerging throughout the dissertation. Specifically, the refined understanding of the HOT theory reveals that there is an important and largely unnoticed separation between the account of phenomenal character we see emerging from the HOT theory and the account of mental state consciousness that is more often the target of philosophical discussions of the HOT theory. Furthermore, despite failing to get as much critical attention, we learn that the HOT theory's account of phenomenal character may be the more important and powerful account of the two offered by the theory. In light of these emerging ideas, the most

important lesson I hope this dissertation provides is that we ought to turn our philosophical attention away from the HOT theory's account of mental state consciousness and toward this theory's rich and nuanced theory of phenomenal character.

References

Block, N. (2011a) "The Higher Order Approach to Consciousness Is Defunct", *Analysis, 71(3)*, p. 419-431.

Block, N. (2011b) "Response to Rosenthal and Weisberg" *Analysis, 71(3)*, p. 443-448.

Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge: Cambridge University Press.

Dretske, F. (1995) *Naturalizing the Mind*, Cambridge, MA: MIT Press.

Gennaro, R. J. (1996) *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*, Amsterdam: John Benjamins Publishing Company.

Hare, B., Call, J., and Tomasello, M. (2001) "Do Chimpanzees Know What Conspecifics Know and Do Not Know?", *Animal Behavior*, *61*, p. 139-151.

Lurz, R. W. (2003) "Advancing the Debate Between HOT and FO Accounts of Consciousness" *Journal of Philosophical Research, 28*, p. 23-44.

Lurz, R. W. (2006) "Conscious Beliefs and Desires: A Same-Order Approach" In Uriah Kriegel and Kenneth Williford (Eds.) *Self-Representational Approaches to Consciousness*, Cambridge, MA: MIT Press, p. 321-351.

Lycan, W. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.

Onishi, K. H. and Baillargeon, R. (2005) "Do 15-Month-Old Infants Understand False Beliefs?", *Science, 308*, p. 255-258.

Perner, J., Leekam, S. R., and Wimmer, H. (1987) "Three-Year-Olds' Difficulty with False Belief: The Case for a Conceptual Deficit" *British Journal of Developmental Psychology, 5*, 125-137.

Povinelli, D.J., Nelson, K.E., and Boysen, S.T. (1990) "Inferences about Guessing and Knowing by Chimpanzees (*Pan troglodytes*)", *Journal of Comparative Psychology*, *104*, p. 203-210.

Premack, D. and Woodruff, G. (1978) "Does the Chimpanzee Have a Theory of Mind?" *The Behavioral and Brain Sciences*, *4*, 515-526.

Rosenthal, D. M. (1986) "Two Concepts of Consciousness", *Philosophical Studies, 491,* p. 329-359.

Rosenthal, D. M. (2002) "A Theory of Consciousness", In Ned Block, Owen Flanagan, and Güven Güzeldere (Eds.) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press, p. 729-753.

Rosenthal, D. M. (2003) "Unity of Consciousness and the Self", *Proceedings of the Aristotelian Society*, *103*, p. 325-352.

Rosenthal, D. M. (2004) "Varieties of Higher-Order Theory", In Rocco J. Gennaro (Ed.) *Higher-Order Theories of Consciousness: An Anthology*, Amsterdam: John Benjamins Publishing Company, p. 17-44.

Rosenthal, D. M. (2005) *Consciousness and Mind*, Oxford: Clarendon Press.

Rosenthal, D. M. (2011) "Exaggerated Reports: Reply to Block" *Analysis, 71(3)*, p. 431-437.

Seager, W. (2004) "A Cold Look at HOT Theory", In R. J. Gennaro (Ed.), *Higher-Order Theories of Consciousness: An Anthology*, Philadelphia, PA: John Benjamins North America, p. 255-275.

Southgate, V., Senju, A., and Csibra, G. (2007) "Action Anticipation Through Attribution of False Belief by 2-Year-Olds", *Psychological Science*, *18(7)*, p. 587-592.

Wimmer, H. and Perner, J. (1983) "Beliefs about Beliefs: Representations and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception" *Cognition*, *13*, p. 103-128.

Paper 1: ***Against the Theory of Mind Objection and the General Use of Theory of Mind Research in Assessing the Higher-Order Thought Theory of Consciousness***

1. Introduction

According to the Higher-Order Thought Theory of Consciousness, a mental state is conscious if and only if it is represented by an appropriate higher-order thought (HOT) to the effect that one is in that particular mental state. This entails, however, that for a person to have any conscious mental states at all, she must have the general capacity to represent mental states in thought. Many see this requirement as a reason to reject the HOT theory. Specifically, objectors argue that very few creatures on earth are capable of representing mental states in thought, and hence that the HOT theory entails that very few creatures on earth are capable of having conscious mental states at all. The counter-intuitiveness of this claim is taken as reason for rejecting the HOT theory as an explanation of mental state consciousness.

I refer to this objection as the *Theory of Mind Objection* because the objectors appeal to the Theory of Mind (TOM) research conducted in Developmental and Comparative Psychology in order to support their claim that few creatures have the capacity to represent mental states in thought. Studies in these fields reveal that most non-human animals and most human infants lack a competence in Theory of Mind, i.e., these subjects are unable to successfully predict or explain someone else's behaviour when doing so requires attributing to that other person certain mental states. As mentioned, objectors then reason from this evidence to the conclusion that those who fail

these tests must lack the capacities the HOT theory requires for mental state consciousness.

Since this objection was first formulated, new evidence has emerged showing that human infants actually *can* pass nonverbal versions of the standard TOM tests. HOT theorists now might appeal to this new evidence and argue that the success of infants on these new nonverbal TOM tests shows that infants do have what the HOT theory requires for mental state consciousness after all. Hence the defender of the HOT theory might thereby create her own TOM argument *against* the TOM objection.

My goal in this discussion is to identify a common assumption held by both sides of this debate and to argue that this assumption is unwarranted. Specifically, in order to take the success or failure of subjects on TOM tests as evidence for or against the HOT theory, one must assume that the requirements of successful performance on TOM tests parallel the very requirements for mental state consciousness that are outlined by the HOT theory. I call this the Parallel Requirements Assumption and I argue that it turns out to be false when assessed in relation to the standard verbal TOM tests (which are taken as support for the TOM objection) and, at the very least, that it turns out to be not clearly true when assessed in relation to the new nonverbal TOM tests (which can be taken as evidence against the TOM objection). Since philosophers on either side of the debate are not justified in holding this assumption, I conclude that both the TOM Objection and the HOT theorist's own argument against the TOM objection must be rejected.

To keep this paper to a manageable length, I limit my inquiry to an assessment of the standard Sally-Anne TOM test paradigm, as well as its new non-verbal counter-parts.

I also restrict my discussion to only the experiments involving typically developing human infants and toddlers, for the most part. Finally, I ignore the possible issue of whether these *other*-directed tests really do measure the *self*-directed ability that HOT theory requires[1] and I restrict this discussion to David Rosenthal's (2002a, 2002b, 2005) Actualist version of the HOT theory of consciousness.[2]

My plan for the paper is as follows: I begin with an outline of the HOT Theory of Consciousness, focusing mainly on points that will be relevant to my later assessment of the two TOM-based arguments. I then introduce the two types of TOM tests and the two TOM-based arguments, one being the TOM objection to HOT theory and the other being the argument on behalf of the HOT theory against the TOM objection. Here I also introduce the Parallel Requirements Assumption that grounds both of these arguments. Next, I identify some important differences between the two forms of TOM tests, which help me in finally articulating the demands of both tests in the language of the HOT theory. I conclude by demonstrating that the demands of these TOM tests are not parallel to the demands of the HOT theory, thus that the Parallel Requirements Assumption is false and that both arguments must be rejected.

---

[1] The problem referred to here is as follows: Most theory of mind tests ask a subject to focus on the behaviour and mental states of *others*, rather than on the subject's own mental states and actions. For example, a subject might be asked to predict what another person will do when that person arrives in the testing room. The HOT Theory, on the other hand, would require for consciousness only that a person be able to pick out *her own* mental states. As Carruthers (2009) argues, however, the ability to pick out one's own mental states might be completely separate from the ability to pick out the mental states of others. The implication would be then that these other-directed TOM tests are not measures of the self-directed skills required by the HOT theory. The fact that the HOT theory and the TOM tests might tap these different skills, and indeed the very theory that these are in fact different skills, are two topics we will not address here. I mention them only to alert the reader.

[2] Though my argument is presented as a defense of Rosenthal's Actualist HOT Theory (Rosenthal 2002a, 2002b, 2005), it is an interesting question whether the TOM objection would be applicable, and if so whether the same response would be available, for other sorts of Higher-Order Theories of Consciousness. I highlight this for the reader, though it is not a question I intend to address here. For examples of other Higher-Order accounts of consciousness see Gennaro (1996), Carruthers (2000) and Lycan (1996).

2. The Higher-Order Thought Theory of Consciousness

To introduce Rosenthal's Higher-Order Thought Theory of Consciousness, it is helpful to begin by looking at some of the assumptions on which the theory is founded. One such assumption is that certain mental states can be conscious at one moment and unconscious at the next, or vice versa. The property a state is said to have when it's conscious and is said to lack when it's unconscious is called *state consciousness*.[3] The HOT theory is a proposal about how best to explain state consciousness.

State consciousness can be contrasted with other sorts of properties we might pick out with the term 'consciousness'. For example, we might say of a person that she is conscious *of* something. Since 'conscious' is here used in a transitive sense – it requires the specification of a direct object – this sort of consciousness is referred to as transitive consciousness, and since it is attributed solely to people or other suitable creatures – we don't say of a mental state that it is conscious *of* something – this sort of consciousness is referred to more formally as *transitive creature consciousness.* We might also say of a person that she is conscious, full-stop, meaning that she is awake and responsive to stimuli as opposed to being unconscious, knocked out, or asleep. Since this sense of

[3] Though HOT theorists most often claim that state consciousness is a relational property (i.e., that it is a property that consists in the intentional relationship between a first-order state and an appropriate HOT about that first-order state, for example, see Rosenthal (2002a)), in his most careful moments Rosenthal says that this characterization is not strictly speaking accurate. Rather, Rosenthal says that state consciousness does consist in *something's* being a state one represents oneself as being in via an appropriate HOT, but he also insists that this *something* can have this property regardless of whether or not there is an actually instantiated first-order state to which we would normally attribute the property of state consciousness. This means that, strictly speaking, we cannot really consider state consciousness to be a property, even if relational, that is attributed to first-order mental states. Instead, it appears Rosenthal intends that this *something* to which we actually attribute state consciousness is a merely *notional* state, a state that is an intentional item, rather than being an actually instantiated mental state. (For more on this more careful account of state consciousness, see Rosenthal (2003, 2011).) Because adjusting our way of speaking to reflect Rosenthal's most careful view makes the debate I'd like to address here unnecessarily complicated, I will follow the HOT theorists' own example and just treat state consciousness as though it is simply a relational property attributed to a mental state. That being said, I do believe that nothing I say in this paper is incompatible with Rosenthal's more careful account of state consciousness, though I leave it for another time to argue for this point.

'conscious' does not require the specification of a direct object, and since it is attributed to a person or other suitable creature here, it is referred to as *intransitive creature consciousness*.

Aside from these distinctions among different senses of 'consciousness', there is also an important distinction we will need to draw between mere 'awareness' and full-blown 'consciousness'.[4]

Thomas Nagel (1974) famously argues that those wishing to explain consciousness must be wary of overlooking the very feature that makes consciousness interesting in the first place. This feature, according to Nagel, is something subjective and experiential, something it's like *for* an organism to be conscious or to have a conscious mental state.[5] This essential, subjective nature of consciousness is also sometimes referred to as the *qualitative* or *phenomenal* character of consciousness or simply as *what-it's-like-ness*. (I will use these terms interchangeably throughout our discussion.)[6] These terms are all meant to capture, for example, the *redness* one experiences when one consciously sees a red tomato or the *painfulness* one feels when one consciously pinches a finger in a door. The lesson from Nagel's argument is that, whatever we may end up identifying with the designation 'consciousness', we must be sure that we are identifying something that essentially involves this what-it's-like-ness. Hence I propose that we accept Nagel's lesson and therefore reserve any use of the term

[4] This distinction is also suggested in Chalmers (1995).
[5] As Nagel writes, "…the fact that an organism has conscious experience *at all* means, basically, that there is something it's like to *be* that organism…something it is like *for* the organism" (Nagel 1974, p. 436, original emphasis).
[6] Note that I intend no theoretical ties to any one particular theory of phenomenal character when I use the term 'phenomenal character' or any of the other terms mentioned here. For example, despite mentioning Block's (2011) term for this aspect of experience (i.e., 'what-it's-like-ness'), I do not intend to have any ties to the specific theory of phenomenal consciousness that Block himself endorses (see Block 1995). Instead, I just mean to identify what Weisberg (2011) calls the "moderate" reading of these terms.

'consciousness' for all and only those phenomena and properties that involve some affiliated what-it's-like-ness.

On the other hand, research has shown time and again that we can have all sorts of *unconscious* mental states, for which there's nothing it's like to be instantiating them but which nonetheless afford us a sort of *mere awareness* of what they represent. A fascinating example of this can be seen in cases of blindsight.

Due to damage to the striate cortex, patients suffering from blindsight develop a pathological 'blind' region in their visual field where they report that they cannot see any stimuli. Forced-choice experiments soon revealed, however, that patients were actually registering information from these blind regions after all. For example, Weiskrantz et al. (1974), tested blindsight patients by placing images of either vertical or horizontal lines in such a way as to fall within the patients' pathological blindspot. The researchers then forced the patients to choose which way the various lines were slanted. Interestingly, though the patients adamantly denied seeing any of the lines, they actually chose the correct orientation a statistically significant amount of the time. This led the researchers to conclude that the patients were registering information about the lines after all, the problem is just that there's nothing it's like for the patients to be 'seeing' the lines.[7]

There are a few important facts to take note of here. First, because these blindsight patients are able to guess the *correct* orientation of the lines a statistically significant amount of the time, it's rather clear that they must be forming visual representations of the lines and also that they do thereby come to have some sort of awareness of those lines. Second, however, since the blindsight patients also adamantly

[7] For more detailed information please see Weiskrantz et al. (1974), Experiment 3.

deny that they can see any of the lines at all, we must conclude that the awareness of the lines that blindsight patients are afforded by their representations is a kind of *unconscious* awareness. It is this sort of unconscious awareness that we must be careful to contrast with the full-blown *consciousness* discussed above.

To that end, as mentioned, we will reserve the term 'consciousness' for all and only those phenomena associated with some kind of what-it's-like-ness, and we will introduce the term 'awareness' for any sort of mental registration of information that has no affiliated what-it's-like-ness. In light of this clarification, we can redraw the distinctions introduced earlier as follows: We will say that if a mental state is state conscious, then there is something it's like for the bearer of that state to be instantiating that mental state at that time.[8] If there's nothing it's like for a person to be in a particular mental state, then that mental state is not state conscious.

We will also say that if a person or other suitable creature is *transitively creature conscious of* some object, then not only must that person be mentally registering information about the object, but also there must be something it's like for that person to be registering that information. On the other hand, if this what-it's-like-ness is missing, yet there is still reason to think the person is representing some object (as is the case in

[8] We should also note that it is unclear whether or not Rosenthal would agree with this assumption. On the one hand, Rosenthal often says that the phenomenal character we experience when our sensations and perceptions are conscious is due to our representing, via HOTs, various special, qualitative properties that only sensory states have. (See, for example, Rosenthal (2004).) It might, then, appear that Rosenthal would not endorse the view that conscious beliefs, say, have something it's like for their bearers to be instantiating them, because beliefs fail to have these qualitative properties. On the other hand, Rosenthal also says, for example, that, "[a] state's being conscious is a matter of mental appearance – of how one's mental life appears to one. …a state is conscious only if one is subjectively aware of oneself as being in that state" (Rosenthal (2011), p. 431). With this sort of description of conscious experience, which makes no appeals to sensory qualities *per se* but rather only to subjective appearances, there seems to be no good reason to deny that there would be a subjective appearance affiliated with being in a conscious belief in addition to there being subjective appearances of being in conscious sensory states. Since Rosenthal's position is unclear and since our discussion is simpler if we adopt the more inclusive thesis, I will assume in this paper that all state conscious mental states have some sort of affiliated what-it's-like-ness.

blindsight, for example), we will say instead the person is *transitively creature aware of* what she is representing.

Finally, we will attribute *intransitive creature consciousness* only when a creature is awake and responsive to stimuli and there's something it's like for the creature to be awake and responsive. We will introduce the term *intransitive creature awareness* to identify what we attribute to a creature that is merely awake and responsive to stimuli, but for whom there is nothing it's like to be awake and responsive. (For example, a philosophical zombie would be intransitively creature aware.) With this terminology now fully clear, we can carry on with our introduction of the HOT theory.

A second assumption on which the HOT theory is based is actually an insight borrowed from the everyday folk: We find it natural to say that if a person is not aware of her mental state in any way, then her mental state is not conscious. This implies, conversely, that a conscious mental state must be a state a person is aware of being in. HOT theorists take this folk intuition to entail that the difference between a mental state when it's state conscious and when it's unconscious lies in its possessor's awareness of it, and so they set out to explain state consciousness in terms of a person's transitive creature awareness of her own mental states. As you might have guessed, HOT theorists say that one comes to have this sort of transitive creature awareness of one's own mental states by forming higher-order thoughts about those mental states. Specifically, HOT theorists argue that a mental state is state conscious if and only if that mental state is represented by an *appropriate* higher-order thought (HOT).

According to Rosenthal, a HOT is *appropriate* when it is noninferential (i.e. it is not the product of any conscious inference or observation), when it is nondispositional

(i.e., when a person actually instantiates the HOT rather than merely having a disposition to form such a HOT), when it is assertoric (i.e., when the propositional attitude of the thought is one of assertion), and when the HOT represents its bearer as being in a particular mental state (for example, only an appropriately formed HOT with the content, roughly, "I believe that all people are equal" will result in the conscious belief that all people are equal). Though these conditions are important components of the theory, they will not play a crucial role in our current discussion, so I will refrain from providing any further explanation of them here.[9]

Notice that the term 'higher-order' is simply meant to highlight the fact that the required thoughts are mental representations of other mental representations. Psychologists also refer to these sorts of HOTs as metarepresentations. Because it will be important for us to keep track of the level of metarepresentation involved in various thoughts throughout our discussion, I will introduce the term 'representational structure' to pick out the level or order of representation that a particular mental state has. For example, we will say that a belief about objects in the world has a *first-order* representational structure because the belief simply represents objects external to the mind rather than representing any other mental states. On the other hand, we will say that a thought about a first-order belief has a *second-order* representational structure because it does involve metarepresentation – it is a mental representation of another mental state (the first-order belief). If one were to form a doubt about a thought about a belief (perhaps someone in the grips of Cartesian skepticism would be moved to do such a thing) then this doubt will have a *third-order* representational structure, because the doubt

[9] For a concise discussion of the conditions that make a HOT appropriate as well as the arguments leading to the HOT theory account of state consciousness presented here see Rosenthal (2002b), especially Section II, "The Hypothesis", pg. 408-411.

is a representation of another mental state (the thought) which itself is also a representation of another mental state (the belief). And so we can keep attributing higher and higher representational orders as the representational structure of a mental state gets more and more complex. To put the central thesis of the HOT theory in this new terminology then, HOT theorists argue that a first-order mental state is state conscious only when it is represented by an appropriate second-order thought.

The explanation so far is an explanation of *basic* state consciousness. *Basic* state consciousness results from the formation of a second-order thought representing oneself as being in a particular first-order state and it is taken to be the sort of state consciousness that we have unreflectively and most often in everyday life. HOT theorists also provide an account of *introspective* state consciousness. *Introspective* state consciousness results from the formation of a third-order thought representing oneself as being in a second-order state and it is taken to involve a more deliberate sort of reflection on our own mental states.[10] Such a third-order thought, for example, might have the content, roughly, "I think that I believe that all people are equal" and the formation of such a thought will result in one's second-order thought becoming state conscious.

There are two interesting things to note about the difference between these two accounts. First, since the HOTs required for basic state consciousness are not, themselves, the objects of any higher-order thoughts, they are not, themselves, state conscious mental states. Rather, only when one forms a third-order thought about one's second-order thought, and hence, only when one has introspective state consciousness, does that second-order thought become state conscious itself. Second, there will be a

[10] For further discussion of the difference between basic and introspective state consciousness see, for example, Rosenthal (2005), Ch. 4.

phenomenological difference between basic and introspective state consciousness. To draw this out, let's look at the HOT theory's account of the phenomenal character of state consciousness.

We agreed earlier that in attributing state consciousness, HOT theorists must be claiming that there comes to be *something it's like for the creature* who instantiates the state conscious mental state. Further, we agreed that when a mental state is not state conscious, there must not be anything it's like for the creature to be instantiating that state. Considering the simpler case first, i.e., *basic* state consciousness, there are two separate representations involved in basic state consciousness, a first-order representation of objects and facts in the world and a second-order representation of oneself as being in that first-order state.[11] This means that there are two options for what one might become conscious of, and hence two options for characterizing what it's like for one, when one's mental state is state conscious: One might become conscious of what one's HOT represents or one might become conscious of what one's first-order state represents.

Some philosophers, for example, Lurz (2006), do take the HOT theorists to be endorsing the former claim but I think it's pretty clear HOT theorists cannot be doing that. After all, the HOTs that afford basic state consciousness are, themselves, unconscious mental states and, in categorizing them as unconscious states, HOT theorists must mean that there is nothing it's like to be instantiating those HOTs. If HOT theorists were to claim that basic state consciousness involves becoming conscious of what one's

[11] Strictly speaking, the HOT theory does not require that both an appropriate HOT as well as a first-order state actually be instantiated in order for there to be basic state consciousness. Instead, Rosenthal's theory allows for the admittedly rare possibility of 'empty HOT' cases, wherein a HOT misrepresents its bearer as being in a particular lower-order state, even though no such lower-order state is instantiated. (This is the sort of possibility that Rosenthal means to account for with his more careful description of state consciousness, as discussed in footnote 3.) As I mentioned earlier, however, we will ignore this complication here, in order to avoid unnecessarily complicating our discussion.

HOT represents, then they would be claiming that there *is* something it's like for one to be instantiating that HOT, hence they would be contradicting their own claim that the HOT is not state conscious. Instead, I propose that HOT theorists are actually making the second claim introduced above, namely, that when one has a state conscious state in the basic sense, one comes to be conscious of what one's *first-order* state represents. Let's formalize this insight with the Phenomenal Character Principle:

> **The Phenomenal Character Principle**: When a mental state, *M*, is represented by an appropriate HOT, and hence when *M* is state conscious, the creature instantiating these states will become transitively creature conscious of what *M* represents.

There's another side to this story as well however. Earlier we learned that when one's mental states are not state conscious, one still can be afforded a mere *awareness* of what one's mental states represent. (We saw this in the case of blindsight for example.) We took this to mean that someone can be differentially responsive to what her unconscious state represents, even though there's no what-it's-like-ness affiliated with her being in that state. We can formalize this point with the Awareness Principle:

> **The Awareness Principle:** Let *M* be any mental state that at times can be state conscious and at other times can be unconscious. When a subject instantiates *M* without also forming an appropriate HOT about being in *M*, *M* will enable the subject to be transitively creature aware of what *M* represents and hence differentially responsive to what *M* represents, but since *M* will not be state conscious, there will not be anything it's like for the subject to be instantiating *M* at that time.

If we take these two principles together now, we see that a full characterization of what it's like for a person who has a basic state conscious mental state is as follows: A person who has a state conscious mental state, in the basic sense, will be transitively creature *conscious* of what her first-order mental state represents and, since she forms a second-order thought that is unconscious, she will also be transitively creature *aware* of

what her second-order thought represents. So, for example, if someone instantiates the belief that all people are equal, and this person also forms an appropriate HOT that she has such a belief, she will become transitively creature conscious of the fact that all people are equal and she will become transitively creature aware of herself as having that belief.

With these two principles we can draw out a full characterization of the phenomenal character of *introspective* state consciousness as well. Since, as we saw, introspective state consciousness involves forming a third-order thought that, in turn, makes it the case that one's second-order thought is state conscious, we can see that a person with an introspectively state conscious mental state will become transitively creature *conscious* of what her second-order thought represents and she will become transitively creature *aware* of what her third-order thought represents. So, in terms of the example above, a person will come to be transitively creature conscious of herself as believing that all people are equal (and thereby also transitively creature conscious of the fact that all people are equal) and she will come to be transitively creature aware of herself as thinking that she believes that all people are equal (since this is what she represents with her unconscious third-order thought). And thus we see the HOT theory's account of what it's like for a person to have state conscious mental states in both the basic and the introspective senses of state consciousness.

There is one final and rather simple point we must draw out about the HOT theory. Since a HOT is simply a thought about a lower-order mental state, so to form a HOT in the first place one generally must be capable of representing mental states in thought. As we will see, this simple point is the catalyst for the TOM objection.

Thus we conclude the introduction of the relevant aspects of the HOT theory. Let's quickly summarize this section by highlighting the main points that will be relevant to our later discussion: (1) The HOT theory dictates that a mental state is state conscious if and only if it is represented by an appropriate HOT to the effect that one is in that very state. Basic state consciousness involves the formation of unconscious second-order thoughts about one's first-order mental states whereas introspective state consciousness involves the formation of unconscious third-order thoughts about one's second-order mental states. (2) One must have the general capacity to represent mental states in thought if one is going to form the HOTs about one's lower-order mental states that are required for state consciousness. (3) When appropriately formed, a HOT about a lower-order state makes its possessor transitively creature *conscious* of what the lower-order state represents. (4) The formation of certain unconscious mental states will enable a creature to become transitively creature *aware* of what those states represent, though there will be no what-it's-like-ness affiliated with those states so long as they fail to be state conscious.

3. Two Theory of Mind Tests, Two Theory of Mind Arguments, One Assumption

Keeping these points about the HOT theory in mind, we are now ready to introduce the two sorts of theory of mind (TOM) tests and the two arguments, one against the HOT theory and one in support of the HOT theory, that might be based on them. We'll begin by discussing the standard TOM tests and the TOM objection to the HOT theory.

3.1 Standard Verbal Theory of Mind Tests

A child is said to have a *theory of mind* when she is able to attribute mental states to herself and others and she is able to explain and/or predict behaviour on the basis of those attributions.[12] In one of the most common tests for these abilities in preschoolers, the Sally-Anne task, a subject is introduced to two puppets, Sally and Anne. The puppets are then used to act out the following story: Each puppet has her own container – Sally has a basket and Anne has a box. Sally has a marble that she plays with for a while and then places in her basket. Sally then leaves the scene. Next, Anne moves Sally's marble to the box (without Sally seeing or knowing this). Sally then returns to play with her marble. At this point in the story the subject is asked, "Where will Sally look for her marble?". The correct answer is that Sally will look in the basket. To answer the question correctly, the subject must understand that Sally would still (falsely) believe that her marble is where she left it and that Sally would act in accordance with her (false) belief. Generally, typically developing children begin to pass this test somewhere between the ages of three and four years old.[13]

Notice that success on this task appears to require that a child be able to pick out mental states and categorize those mental states in terms of their distinctively mental properties and typical causal interactions. For example, in order to predict that Sally will search in the basket, a child in this experiment must pick out Sally's belief about the location of the marble and then figure out that Sally's belief will influence where Sally searches for the marble. It's clear that if a child is to achieve any of this however, she

[12] Premack and Woodruff (1978) introduced the term 'theory of mind'. This ability is also sometimes referred to as the ability to mentalize or to mindread.
[13] For early studies demonstrating these results see Wimmer and Perner (1983) and Perner, Leekam, and Wimmer, (1987).

must be able to represent mental states in thought in the first place. Hence there appears to be a parallel in the cognitive requirements for success on these TOM tests and the requirements for HOT formation as set out by the HOT theory.[14] This presumed parallel in requirements plays a central role in both of the arguments we will consider so let's represent it formally here:

> **The Parallel Requirements (PR) Assumption**: The general capacity to represent mental states in thought is both necessary for a subject to perform successfully on TOM tests and necessary for a subject to form the kinds of higher-order thoughts the HOT theory requires for basic state consciousness.

Let's now see how this PR Assumption is implicated in the TOM Objection.

3.2 The Theory of Mind Objection to the Higher-Order Thought Theory of Consciousness

Proponents of the TOM objection[15] begin by noting that surprisingly few subjects pass TOM tests. For example, as we just saw, typically developing children seem incapable of passing standard verbal TOM tests until around the age of 3 to 4 years.

---

[14] This point about shared requirements is more often phrased in terms of a parallel in *conceptual requirements*. Since the HOT theory requires one to from HOTs about one's lower-order mental states in order to make those lower-order states conscious, and since the formation of propositional thoughts like these is generally taken to involve the activation of concepts for the objects or facts represented, therefore in order to have any HOTs in the first place it would appear that a creature must activate a concept picking out the particular mental state it is representing with its HOT (for example, BELIEF THAT P). On the other hand, the very skills that seem to lead to success on TOM tasks – picking out mental states and categorizing them in terms of their distinctively mental properties and typical causal interactions – appear to be the very same sorts of skills that the possession of concepts for our mental states would afford. Since the formation of HOTs and one's successful performance on TOM tests both appear to require that a person possess and activate mental state concepts, we see, again, a parallel in requisite abilities similar to the one outlined above – the possession and activation of mental state concepts appears to be necessary both for a creature to form the kinds of thoughts HOT theory requires for basic state consciousness and for a creature to perform successfully on TOM tests. I choose not to frame the debate in terms of concept possession however, and instead frame it in terms of *a creatures capacity to represent mental states in thought*, because there is still a lot of disagreement among philosophers about the precise nature of concepts and about the precise concepts the HOT theory would require for state consciousness (for example, see Rosenthal, (2000), p. 279). By instead framing the debate in terms of a creature's capacity to represent mental states in thought I believe we actually can make progress in this debate while avoiding the messy issues associated with concept possession.

[15] For examples of this objection in the literature see Dretske (1995), Ch. 4 or Seager (2004).

Furthermore, and despite sometimes showing quite sophisticated cognitive abilities in other areas, most high-functioning people with Autism Spectrum Disorders[16] fail to show TOM abilities until they achieve a verbal mental age of about 9 years[17], if they ever show these abilities at all. Finally, most non-human animals, including the great apes, appear incapable of passing these tests either.[18] Objectors argue that those failing these tests must also lack the capacity to represent mental states in thought that the HOT theory requires for basic state consciousness.

Notice that, in counting a subject's failure on these tests as evidence that the subject lacks the mental capacities the HOT theory requires, the objector must be relying on the PR Assumption outlined above, namely, that the general capacity to represent mental states in thought is both necessary for a subject to perform successfully on TOM tests and necessary for a subject to form the kinds of higher-order thoughts the HOT theory requires for basic state consciousness. Only with this assumption can the objector

---

[16] The term 'Autism Spectrum Disorder' covers a spectrum of developmental disorders that are characterized primarily by the presence of all three of the following behavioural measures: (1) impairments in social interaction, (2) impairments in communication, and (3) unusually restricted behaviours and interests. Autism Spectrum Disorders are usually diagnosed in early childhood and the severity of symptoms spans from quite mild to quite severe, with the more severe cases often accompanied by other disorders such as epilepsy and learning disabilities. Interestingly, in light of our topic, a popular hypothesis of the underlying cause of Autism Spectrum Disorders is that they stem from a breakdown in the mechanisms realizing Theory of Mind abilities. For more on this *Mindblindness* hypothesis see Baron-Cohen (1995) and for an excellent general overview of Autism Spectrum Disorders see Frith (2003).

[17] Happé (1995).

[18] For example, Povinelli, Nelson, and Boysen (1990) found that chimpanzees were unable to identify that they should ask for food from an experimenter who watched where the food was placed over another experimenter who could not see where the food was placed because she had a bucket over her head. The researchers take this as evidence that chimpanzees are unable to take the visual perspective of others. In direct contrast to these findings however, Hare, Call, and Tomasello (2001), have found evidence that chimps can adjust their food searching behaviour in relation to whether or not a more dominant chimp has or has not seen where desirable food was hidden. Specifically, when the dominant chimp has not seen the placement of the desirable food the subordinate chimp will retrieve that food but when the dominant chimp has seen the food placement the subordinate chimp will stay away from the food location. Though researchers continue to debate about the theory of mind skills chimps may or may not have, there has yet to be any uncontroversial evidence of false belief understanding in this population.

argue from a subject's failure on these tests to the conclusion that the subject fails to have the metarepresentational abilities the HOT theory requires.

The rest of the argument is rather straightforward. Since so many subjects fail these tests, and since failure on these tests is taken as evidence that these subjects lack what the HOT theory requires for basic state consciousness, objectors reason that HOT theorists are forced to conclude that those failing these tests are incapable of basic state consciousness. Since this conclusion is taken to be counter-intuitive, the fact that the HOT theory leads to such a conclusion is counted as evidence that the HOT theory must be mistaken and should therefore be rejected as an explanation of basic state consciousness.

This problem is not just a theoretical problem for the HOT theory; it also has some quite important consequences. For example, philosophers argue that this problem has surprising implications for the issue of animal rights.[19] If the objector is correct and TOM tests show that animals do not have the cognitive capacities required for basic state consciousness, then HOT theorist would be forced to say that animals cannot *consciously* experience suffering. Hence the HOT theory would be calling into question the view that animals are sentient beings. If animals are not sentient however, then their interests might no longer count in moral deliberation.[20] Thus the HOT theory might entail not only that animals fail to have any state conscious mental states, but also that animals deserve no moral consideration.

---

[19] For example, see the exchange between Carruthers (1989, 2000) and Gennaro (1993, 1996, 2004).
[20] Though Carruthers (1989) originally endorsed this sort of argument, he has since changed his mind. Carruthers (2000) now argues that moral concerns can be grounded in first-order desire frustration so higher-order thoughts and thus *conscious* desire frustration is no longer seen to be necessary.

Though I do not intend to follow up on this point here, notice that this issue is even more pressing once we expand the argument to include the young children and people with Autism Spectrum Disorders who also fail these TOM tests. Since these populations also fail TOM tests, they also appear to lack the capacities the HOT theory requires, so they also would lack conscious suffering. Thus the HOT theory might entail that children and people with Autism Spectrum Disorders are not owed moral consideration either. Surely this is a consequence that HOT theorists should try to avoid, so the TOM objection really is a pressing concern.

Again then, the logic of the objection is as follows: Since, according to the PR Assumption, the capacity to represent mental states in thought is necessary both for the formation of the HOTs required by the HOT theory for basic state consciousness and for a subject's successful performance on standard verbal TOM tests, objectors believe that one's performance on these tests can indicate whether or not one has the very capacities the HOT theory requires for basic state consciousness. Since it turns out that so many subjects fail these TOM tests, HOT theorists seem forced to deny that these subjects are capable of having any state conscious mental states whatsoever. Objectors find this conclusion counter-intuitive and instead argue that the fact that the HOT theory leads to such a conclusion is evidence that the theory must be mistaken. The objectors thus take themselves to have shown that the HOT theory ought to be rejected as an explanation of state consciousness.

With this understanding of the TOM objection to the HOT theory, let's turn now to a discussion of the new non-verbal TOM tests and the argument in favour of the HOT theory that might be based on them.

3.3 New Non-verbal Theory of Mind Tests

Onishi and Baillargeon (2005) used the violation-of-expectation paradigm (a paradigm based on the assumption that infants will look longer at surprising events) to test whether infants as young as 15 months would expect an actor to behave in accord with her false beliefs. The experiment is set up similarly to the traditional Sally-Anne test, though the whole procedure is carried out nonverbally. In the belief-induction phase, infants watch as an actor places a toy in one of two boxes. The scenario then progresses in one of four ways, in order to induce in the actor either a true belief or a false belief about the toy's location: Either the toy changes locations while the actor watches (resulting in a true belief), or it changes locations while the actor does not watch (resulting in a false belief), or it remains in the same location while the actor watches (resulting in a true belief), or it changes locations while the actor watches and then changes back to the original location while the actor does not watch (resulting in a false belief). The infants then receive the test trial, where their looking times are recorded as they watch the actor reach for the toy in one of the two locations. Onishi and Baillargeon found that infants were always surprised (i.e., looked for significantly longer) when the actor failed to act in accord with her beliefs, regardless of whether those beliefs happened to be true or false. As the authors write, "[t]hese results suggest that 15-month-old infants already possess (at least in a rudimentary and implicit form) a representational theory of mind: They realize that others act on the basis of their beliefs and that these beliefs are representations that may or may not mirror reality" (Onishi and Baillargeon (2005), p. 257).

Southgate, Senju, and Csibra (2007) conducted a similar study in which a toddler and a confederate first watch as a toy is placed in one of two boxes then, while the confederate is distracted, the toy's location changes. Finally, a signal is presented, which indicates to the toddler that the confederate is about to search for the toy. Using an anticipatory looking paradigm this time (which measures where subjects look in anticipation of the actor's action, before the actor reaches for the toy), Southgate et al. measured the direction of the first eye saccade the infant made when the signal was presented as well as the amount of time the infant spent looking at the correct and incorrect locations in the brief pause between the signal and the actor's action. They also found statistically significant results, this time indicating that 2-year-olds looked first and looked for longer overall at the false-belief-target location. As Southgate et al. write, their findings "strongly suggest that 25-month-old infants correctly attribute a false belief to another person and anticipate that person's behaviour in accord with this false belief" (Southgate et al. (2007), p. 590).

3.4 The Higher-Order Thought Theory's Theory of Mind Argument *Against* the Theory of Mind Objection

In light of this new evidence, an argument on behalf of the HOT theory might be formed that, in a way, mirrors the structure of the original TOM Objection. Specifically, if HOT theorists were to adopt the PR Assumption themselves they could reason as follows: Since, according to the PR Assumption, the capacity to represent mental states in thought is necessary both for successful performance on TOM tests as well as for the formation of the HOTs needed for basic state consciousness, an infant's success on these new non-verbal TOM tests is evidence that these infants do have what it takes to have

basic state conscious mental states after all. Since HOT theorists no longer must deny that infants are capable of having state conscious mental states, the HOT theory does not appear to have the sorts of counter-intuitive implications that the objector identifies, and hence there is no longer any reason to reject the theory.

So here we see how philosophers on either side of the debate can use the TOM test evidence to support their differing arguments. Proponents of the TOM objection can appeal to a subject's *failure* on verbal TOM test as evidence that the subject does *not* have the capacities that the HOT theory requires, whereas proponents of the HOT theory can appeal to a subject's *success* on nonverbal TOM tests as evidence that the subject *does* have the metarepresentational capacities the HOT theory requires for basic state consciousness.

It's also clear that, in taking a subject's performance on TOM tests as evidence for either side, both sides of the debate must be making the same PR Assumption, namely, that the very same capacity to represent mental states in thought is necessary both for the formation of the thoughts the HOT theory requires for basic state consciousness and for successful performance on these TOM tests. In holding the very same assumption, however, both sides are also vulnerable to the very same criticism. Specifically, if it turns out that the PR Assumption is unwarranted, then both arguments will have to be rejected. I believe this does turn out to be the case but, interestingly, that the PR Assumption fails in each case for different reasons. In order to draw this out, we'll have to determine the actual requirements for successful performance in each of the TOM test paradigms. Let's begin that discussion by identifying some of the important differences

between the two types of TOM tests, in order to draw out some facts that will help us in assessing what each paradigm really demands of its subjects.

4. Some Differences Between the Verbal and Non-Verbal Theory of Mind Test Paradigms

As noted, the key to my argument is an assessment of the actual requirements each sort of TOM test has for successful performance. In order to properly identify those requirements, however, we must first identify a few important differences between these verbal and nonverbal TOM test paradigms.

One obvious difference between the new nonverbal TOM tests and the standard verbal TOM tests is that the new tests do not involve any verbal communication. Specifically, in the new tests neither the narrative, nor the test questions, nor the subjects' responses involve verbal behaviour, whereas in the Standard Sally-Anne test children are presented with a verbal narrative, are directly asked the test question (for example, 'Where will Sally look for her marble?'), and are encouraged (though not necessarily required) to respond verbally (for example, by saying, "in the basket").[21]

This shift is important because it eliminates verbal competence as a confounding factor in the success or failure of a child on the false belief task. After all, the ages at which substantial development in TOM abilities seems to occur are the very same ages at which subjects are developing their ability to use language as well. So, if a researcher's only measure of TOM abilities relies on verbal competence, children might fail simply

[21] Though a verbal response may not always be required, explicit communicative behaviour is always required. For example, Wimmer and Perner (1983) allow children simply to point to a location in response to verbal test questions about where the story character will search.

because they lack that competence rather than failing because they lack the capacity to represent mental states in thought.[22]

Another important distinction between these tests was hinted at in Onishi and Baillargeon's concluding remarks about their experiment. Recall, they write that their "results suggest that 15-month-old infants already possess (at least in a rudimentary and *implicit* form) a representational theory of mind…" (Onishi and Baillargeon (2005), p. 257, emphasis added). As we see in this quote, there's a tendency in the field to refer to the old standard tests as measures of an *explicit* knowledge of the mind and the new nonverbal tests as measures of an *implicit* understanding of the mind. Though there's little formal reflection on what these terms are meant to pick out in this field, the general consensus seems to be that the explicit knowledge is later to develop and it is conscious, reportable, and perhaps consciously controlled, whereas the implicit knowledge is earlier to develop and it is unconscious, nonverbalizable, and can influence behaviour without any conscious awareness of it doing so.[23]

---

[22] In fact, this is a central issue in debates over the true reason *why* subjects fail standard theory of mind tests. Some researchers endorse Competence accounts of failure and argue that subjects fail TOM tests simply because they lack the knowledge of the mind the test is designed to measure. (For example, see Wellman (1991). Note that Wellman suggests the shift in knowledge actually occurs between the ages of 2 and 3 years old for typically developing children, rather than between 3 and 4 years old, since he found evidence that 3 year-olds can explain, though they cannot yet predict, other's actions based on attributions of false-beliefs. To keep our discussion as simple as possible however, I will continue to follow the majority in saying that the standard age that a shift is proposed to occur is between 3 and 4 years.) On the other hand, some researchers endorse Performance accounts of failure and argue that TOM tests rely not only on a creature's knowledge of the mind, but also on other aspects of cognition such as linguistic competence, executive control (a term which, roughly speaking, covers the many cognitive functions that allow a creature to plan and initiate goal-directed behaviour), and the ability to handle sufficient computational complexity. Those on this side of the debate argue that a breakdown in any of these other areas would prevent a subject from performing successfully on TOM tests, *even if* she had the knowledge the test requires. (For example, see Fodor (1992).) A member of this camp would likely raise the worry introduced in the text.

[23] To see further support for this assessment of the field, consider the following quotes from researchers actively engaged in studying infants' TOM abilities: "Appealing to implicit knowledge when infants show correct looking is, in fact, a very popular option in the infancy field…Yet there is typically no attempt to define what is meant by the term "implicit" except "earlier developing" (Ruffman et al. (2001), p. 202); "The distinction between implicit and explicit knowledge plays an important role not only in cognitive

This way of understanding the distinction between an implicit understanding of the mind and an explicit knowledge of the mind is apparent, for example, in a study by Ruffman, Garnham, Import, and Connolly (2001). These researchers set out to investigate whether the anticipatory eye gaze measure, which was subsequently used by Southgate et al. (2007), taps into an implicit understanding rather than an explicit knowledge of the mind by testing whether or not children are conscious of the information that their eye gaze expresses. Since the findings of this experiment can, in addition to drawing out this distinction between implicit and explicit knowledge, also help us in identifying some of the differing demands of the two sorts of TOM tests, we'll take a moment to discuss some of the details of this experiment here.

To measure whether a child is or is not conscious of the information expressed by her eye gaze, Ruffman et al. (2001) use a betting protocol wherein children are asked to bet, by placing differing amounts of counters, on each of the predictions they make throughout the experiment. The experimenters argue that the betting forces a subject to measure her certainty in her own answer and that the task of assessing one's own certainty is a task that requires *conscious* processing. The same is true, they note, when subjects are asked to answer verbal questions; conscious processing is also required. Furthermore, and unsurprisingly, the information that a subject consciously processes must itself be made conscious in order to be processed consciously. So, for example, a subject's predictions and assessments of her confidence in those predictions both would need to be conscious in order for her to place bets on her predictions. Or, similarly, a subject's thought about where Sally believes her marble to be would need to be a

---

development but in cognitive science at large, despite the fact that no agreed meaning of this distinction has yet emerged. Our use is primarily descriptive and intuitive" (Clements and Perner (1994), p. 377).

conscious thought in order for the subject to reply to verbal questions about Sally's belief. So the experimenters reason that by using this betting protocol and by assessing whether or not a subject's eye gaze behaviour matches up with her betting behaviour, they can discover whether the information expressed by a subject's eye gaze plays a role in the bets she places. In this way the researchers believe they can determine whether the information expressed by a subject's eye gaze is conscious and explicit or is unconscious and implicit.[24]

Before testing theory of mind knowledge, however, Ruffman et al. first had to ensure that the betting procedure would measure even slight shifts in a child's confidence in her answers. To determine whether this was the case, Ruffman et al. included two important control conditions in their experiment. In both of these control conditions children were presented with an apparatus consisting of two slides. One slide was red and would allow only red squares to slide down, the other was green and would allow only green balls to slide down. Children were then shown a bag with some amount of red squares and/or green balls and were asked to predict which of the two slides an object from the bag would emerge from. They were also asked to bet on their predictions.

In one of these control conditions the bag contained only red squares. In this case, the correct answer would be to predict that an object from the bag would come out of the red slide. Furthermore, given the certainty of the outcome, it was predicted that children should bet on their answer with confidence. In the second control scenario the bag contains nine red squares and one green ball, thus introducing a slight shift in the probability that an answer of 'red slide' is correct. The researchers reasoned that, if the

[24] Since the points discussed here are relatively central to my argument, I direct those interested in following up on these points to Ruffman et al. (2001), p. 203.

betting procedure is a subtle measure of confidence, then this shift in probability should be reflected in the children's betting behaviour as well, since the children can no longer be certain that their answer of 'red slide' is correct.

Indeed subjects did show this pattern of betting responses. When presented the bag of only red objects, the children placed a mean of nine to ten of their total of ten counters at the dominant (red) location, indicating that the children bet confidently on their predictions. On the other hand, when presented the bag of nine red objects and one green object children only bet a mean of five to six of their ten counters at the dominant (red) location, indicating that the children adjusted their bets significantly to reflect even this slight drop in confidence in their predictions.[25] From this evidence the researchers concluded that the betting procedure was in fact a subtle measure of even slight shifts in confidence.

Having established that the betting measure was sensitive enough to detect even slight shifts in confidence, the researchers then presented their subjects – typically developing 3-5 yr olds – with a false belief task similar to the Sally-Anne test. Once again, children are introduced to a character, Ed, who has a toy that he places in one of two locations. He then leaves the scene to take a nap while another character, Katy, enters the scene and moves the toy to the other of the two locations. Katy then leaves the

[25] In addition to these controls, experimenters also ran an ambiguous condition task, to rule out the possibility that children were just matching their bets to the proportions of red or green objects in the bag. In this ambiguous condition, there was only one object in the bag but children could not see it; instead they were just told that it could be red or it could be green. Ruffman et al. (2001) reasoned that if children were just matching bets to objects then they would bet with greater confidence on whichever colour slide they guessed the object to be affiliated with, whereas if they were betting based on actual confidence in their answers they would spilt their bets between the two sides. Researcher found that subjects "were significantly more likely to spread their [betting] counters at more than one location in the ambiguous task relative to the 10-0 [task]. … Children's tendency to spread their counters in the ambiguous task and not spread them in the 10-0 [task]…is consistent with the idea that their betting on all tasks is based on certainty about probabilities" Ruffman et al. (2001), p. 214.

scene as well. Ed then wakes up and wants to play with his toy so he re-enters the scene. At this point, the experimenter wonders aloud where Ed will look for his toy and the children's anticipatory eye gaze is measured. Finally, the children are asked directly where they think Ed will look for his toy and they are also asked to bet on their predictions by placing their counters next to one or both of the possible search locations.

Interestingly, Ruffman et al. found that some children initially look at the correct location (where Ed falsely believes the toy to be) but subsequently answer the verbal question incorrectly (by saying Ed will search where the toy is now actually located), replicating the findings of Clements and Perner (1994) that there is a stage in development where children's performance is split between the eye gaze and the verbal response measures. Even more interestingly however, Ruffman et al. also found that the youngest children who show this pattern of split responding (mean age of 3.4 years) do not seem to be *conscious* of the understanding of the mind that is conveyed by their correct anticipatory looking behaviour. This fact was evidenced by the failure of these young split-responders to take into account the knowledge expressed by their (correct) eye gaze when assessing their confidence in their (incorrect) verbal responses.[26]

---

[26] The following is a summary of the full results of this study: The researchers first focused their analysis on only those children who passed the eye gaze measure. They then split that group of eye gaze passers into four smaller groups. First they separated those children who passed the eye gaze measure but failed the explicit (verbal and betting) measures from those who passed all measures. Then they separated each of these groups in half to get a group of younger children and older children within each group. So they end up with four comparison groups: younger children who pass eye gaze but fail explicit measures (mean age 3.40 years), older children who pass eye gaze but fail explicit measures (mean age 4.09 years), younger passers of both measures (mean age 3.59 years) and older passers of both measures (mean age 4.46 years). All groups bet with significantly less confidence in the 9:1 task than the 10:0 task, indicating they were all sensitive to shifts in confidence. The rest of the results breakdown as follows: Older split-responders bet on their (incorrect) verbal reply to the false belief task with significantly *less* confidence than their bets in the 10:0 scenario (showing that they did (consciously) register some uncertainty about their reply) but with significantly *more* confidence than their bets in the 9:1 scenario (showing that the shift in confidence was less than the shift induced by the 9:1 scenario). Younger children who passed both measures bet on their (correct) verbal reply to the false belief task with significantly *less* confidence than their bets in the 10:0 scenario (showing that they too (consciously) registered some uncertainty about their reply) but showed no

To explain this conclusion a little further, researchers found that, despite looking to the correct location, these young split-responders bet just as confidently on their incorrect verbal reply as they did when they were betting in the control scenarios where the outcome was guaranteed (i.e., the 10 red:0 green scenarios). These young split-responders also showed significantly *more* confidence in their incorrect verbal reply than they did when betting on the control scenarios where the outcome was slightly less than guaranteed (the 9 red:1 green scenarios).[27] Since the shift in betting behaviour measured in these 9:1 control scenarios demonstrates that these children can be sensitive to even subtle changes in their own confidence, and since the youngest children showing these split responses do not show a similar shift in confidence when betting on their incorrect verbal responses to the Sally-Anne-style task, Ruffman et al. conclude that the understanding of the mind expressed by these children's correct eye gaze responses must not be a conscious understanding, or else it would influence the subjects' betting behaviour. Hence the researchers conclude that the eye gaze measure is actually measuring an unconscious and implicit understanding of the mind in these young

---

significant difference in betting relative to the 9:1 scenario (showing the shift in confidence was comparable to the one induced in the 9:1 scenario). This group was the least confident in their false-belief answers. Finally, older children who passed both measures showed no significant difference between their betting on their (correct) verbal reply to the false belief task and their bets in the 10:0 scenario (showing a similar certainty in both answers), but bet on their (correct) verbal reply to the false belief task with significantly *more* confidence than their bets in the 9:1 scenario (showing, again that they were relatively certain about their reply in the false-belief task). (Note, the researchers report in the text that the older passers "were more certain [in their false-belief answers] than on the 9-1 task" (Ruffman et al, (2001), p. 211) and this statement is similar to the statements they made when reporting other statistically significant results. That being said, the researchers do not report the difference in betting here as statistically significant in one of the tables summarizing their data. I assume this was an oversight on their part but I mention this fact for the reader.)

[27] More specifically, 94% of these young split responders bet all their counters on the (incorrect) location identified with their verbal reply, despite looking at the correct location initially. On the other hand, 83% of these young split responders showed a sensitivity to shifts in confidence by betting at least some counters on the non-dominant (green) location in the 9 red:1 green control scenario.

children.[28] On the other hand, since conscious processing is required both in order to bet on one's predictions and in order to reply to verbal questioning, the researchers argue that the knowledge of the mind that is conveyed by these other measures must be both conscious and explicit.

Adding this lesson to the one identified earlier, we've now learned that the two types of TOM tests have the following differences: Standard Tom tests are primarily verbal and they tap into a subject's explicit, and therefore conscious, knowledge of the mind. On the other hand, nonverbal TOM tests are not reliant on a subject's verbal abilities and they tap a subject's implicit, and therefore unconscious, understanding of the mind. Here we see then, that there are clearly some significant differences between these tests, and hence we can expect the requirements for success on each test to differ significantly as well. Let's now turn to an analysis of those requirements and an assessment of whether the PR Assumption is justified in relation to either test paradigm.

5. The Requirements for Success on Theory of Mind Tests and the Fate of the Parallel Requirements Assumption

To tease out the demands that these TOM tests place on their subjects we will focus on the *representational structure* of the mental states that a subject must form if she is to be successful on the test. Again, by 'representational structure' I mean to identify the level or order of representation that a particular mental state involves. For example, first-order mental states are representations of non-mental objects or facts in the world, second-order mental states are representations of first-order mental states, and so forth.

[28] As Ruffman et al. write, "[o]ur finding that younger failers were more certain on the false belief task than on the 9-1 task is consistent with the idea that despite looking to the [correct] left-hand location, such children are not conscious that the story character might return there," (Ruffman et al. (2001), p. 211).

The reason for focusing on the representational structure of the thoughts required by these tests should be quite clear: Both arguments are based on the PR Assumption that the capacity to represent mental states in thought is both necessary for successful performance on TOM tests and necessary for the formation of the HOTs required by the HOT theory for basic state consciousness. The target of these arguments is the HOT theory's account of *basic* state consciousness. As we learned earlier, the HOT theory says that what is required for basic state consciousness is the formation of an unconscious appropriate *second-order* thought (about oneself as being in a particular first-order state). Hence the capacity that is required by the HOT theory for basic state consciousness is precisely the capacity to form these second-order thoughts. In order for the PR Assumption to be justified then, it must be the case that successful performance on TOM tests also requires precisely the same capacity to form second-order thoughts. So, by investigating the representational structure of the thoughts required for success on each of these TOM tests, we can thereby assess the soundness of the PR Assumption and determine whether either of the two arguments should be accepted. To that end, let's now look at the demands of each of the two types of TOM tests in turn.

5.1 The Demands of the Verbal False Belief Task

I refer the reader to Table 1, which outlines the representational structure of various schematically depicted thoughts that we will be discussing throughout this analysis.

To begin, notice that the first row represents the belief formed by the fictional character, Sally, in the Sally-Anne tests. Sally believes that the marble is in the basket.

Table 1. *The representational structure of thoughts relevant to the verbal false belief task.*

	Mental State in Question	**Schematic of Representation**	**Order of Representational Structure**
1	Sally's Belief	B <that the marble is in the basket>	1st
2	Subject's Representation of Sally's Belief	T < Sally Believes <that the marble is in the basket> >	2nd
3	HOT Required for *Basic* State Consciousness	T < I Believe <that all people are equal> >	2nd
4	Thought Required for Successful Performance on (Explicit) Verbal TOM Test	T < I Think < that Sally Believes <that the marble is in the basket> > >	3rd
5	HOT Required for *Introspective* State Consciousness	T < I Think < that I Believe <that all people are equal> > >	3rd

As we can see, Sally's belief has a first-order representational structure because it represents non-mental items and facts, it does not represent any other mental states.

Researchers believe that subjects of the verbal false-belief test must represent Sally's belief in order to pass the test, perhaps by forming a thought as represented in the second row of Table 1, namely, the thought that Sally believes that the marble is in the basket. This thought clearly has a second-order representational structure because it is a thought about another mental state (Sally's belief).

Notice that this thought obviously is not the kind of second-order thought that would enable state consciousness of any kind, as it fails to meet at least one of

Rosenthal's conditions on the appropriateness of HOTs, namely, that appropriate HOTs must be representations of *oneself* as being in a particular mental state. Hence, as one would expect, the HOT theory does *not* predict that, in forming a thought like the one in row 2, a subject will come to have *Sally's* belief state consciously.

In the third row we see represented a HOT that, if formed appropriately, would provide for basic state consciousness. Since this thought is a representation of another mental representation (one's first-order belief that all people are equal), this thought also has a second-order representational structure.

Once we compare the thoughts in rows 2 and 3, we see the similarity in representational structure that seems to ground the PR Assumption. If a thought like the one is row 2 is all that's required for successful performance on these tests, and if the representational structure of this thought does mirror the representational structure of the HOTs required for basic state consciousness, then these tests would seem to measure the very capacity to represent mental states in thought that the HOT theory requires for basic state consciousness. The question now is whether a thought like the one is row 2 is all that's required for successful performance on these tests.

We learned from the Ruffman et al. (2001) study that these verbal tests tap into a subject's *explicit* knowledge and that answering verbal questions is a task that requires the subject to engage in *conscious* processing. Both of these facts entail that the subject's knowledge of Sally's belief must be knowledge of which the subject is *conscious*. Furthermore, as we learned from the Phenomenal Character Principle, the HOT theory dictates that a person only comes to be conscious of what her mental states represent when those mental states themselves are represented by appropriate HOTs. Putting these

pieces together, since a subject must be conscious of what she thinks about Sally's belief, and since a subject can only be conscious of what she thinks if her thought itself is state conscious, we can see that, according to the HOT theory, a subject actually would need to form a further HOT about her thought about Sally's belief in order to respond to the verbal questions and pass these tests successfully. Such a thought is represented in row 4.

Now this point is important: Notice that this thought in row 4, the thought that a subject actually must form in order to pass the verbal Sally-Anne test, is a thought that has a *third-order* representational structure. In forming this thought the subject is representing both Sally's belief as well as the subject's own thought that Sally has that belief, hence the subject is actually forced to do something structurally more complex that what HOT theory requires for basic state consciousness.

From here we can draw two conclusions: First, we can conclude that the PR Assumption is actually unjustified in this case. Since these verbal TOM tests require more of a subject than the mere formation of a second-order thought, they require more than the HOT theory requires for *basic* state consciousness. Hence there is no parallel in requirements in these cases. This means that a subject's failure on these verbal TOM tests cannot indicate that the subject lacks the minimal capacity to represent mental states in thought that HOT theory requires for *basic* state consciousness, and so the objection fails.

Second, since the thought actually required for success on these verbal TOM tests is structurally similar to the thoughts required for *introspective* state consciousness (compare rows 4 and 5 in Table 1 for example), one might be tempted to save the

objection by arguing that a subject's failure on these tests instead indicates that the subject lacks the mental capacities necessary for *introspective* state consciousness and hence provides us with a new reason to reject the HOT theory. I think this new argument would be too quick though.

The original objection claimed that the HOT theory was forced toward a counter-intuitive conclusion, namely, that typically developing humans beyond the age of 4 are essentially the only creatures on earth who have any conscious mental states whatsoever. It was the counter-intuitiveness of this conclusion that provided the force to reject the HOT theory. According to the new version of the objection though, the conclusion HOT theorists might be forced to draw is as follows: Typically developing humans beyond the age of 4 are essentially the only creatures on earth capable of *introspective* state consciousness. The problem is that this conclusion is not as likely to be deemed counter-intuitive any more. Furthermore, if we do find that the majority's intuitions are not in conflict with this conclusion (as I suspect that we will), then there would no longer be any impetus to reject the HOT theory. So, because the TOM Objection relies on the counter-intuitiveness of the claims HOT theory is forced to make, and because this new claim does not seem to be as likely to be deemed counter-intuitive, it would appear that the TOM Objection could not be saved by repurposing it as an argument about *introspective* state consciousness. This is our second conclusion.

To summarize, successful performance on verbal TOM tests actually requires a subject to form thoughts with a third-order representational structure, which mirrors the structure of the HOTs required for *introspective* but not *basic* state consciousness. In light of this fact, we must conclude that the PR Assumption fails to be warranted and that

a subject's failure on verbal TOM tests should not be taken as evidence that the subject lacks the ability to form the second-order representations the HOT theory requires for *basic* state consciousness. Since it was only the claim that very few creatures were capable of *basic* state consciousness that seemed to be counterintuitive, the wide-spread failure of essentially all but typically developing humans older than 4 years on these tests can no longer be taken as evidence against the HOT theory. Hence, the TOM objection can be rejected.[29]

5.2 The Demands of the Non-Verbal False Belief Task

Let us now assess the demands that the nonverbal TOM tests place on their subjects and ask whether the PR Assumption might be warranted in the HOT theorist's own TOM argument *against* the TOM objection. This time I refer the reader to Table 2, which again outlines the representational structure of various thoughts that we will discuss throughout this analysis.

As shown in row 1 of Table 2, the actor in these nonverbal TOM tests will form a belief about the toy's location, just as Sally formed a belief about her marble's location in the verbal Sally-Anne test. The belief formed by the actor here will also have a first-order representational structure because it is a representation of non-mental items and facts, not a representation of any other mental states.

Though researchers are in less agreement here (as we'll discuss shortly), let's first assume that successful performance on these nonverbal TOM tests also requires that an infant represent the actor's belief about where the toy is located. Under this assumption we see that, as was the case with the verbal TOM tests, the subject would need to form a

[29] Note that the *success* of subjects on these verbal TOM tests would appear to be a clear indication that subjects do have what it takes (and more) to have basic state consciousness. I note this point for the reader but, as it does not relate to any of the arguments we are assessing, I will discuss it no further here.

Table 2. *The representational structure of thoughts relevant to the non-verbal false belief task.*

	Mental State in Question	**Schematic of Representation**	**Order of Representational Structure**
1	Actor's Belief	B <that the toy is in the green box>	1st
2	Thought *Possibly* Required for Successful Performance on (Implicit) Nonverbal TOM Test	T < Actor Believes <that the toy is in the green box> >	2nd
3	HOT Required for *Basic* State Consciousness	T < I Believe <that all people are equal> >	2nd
4	Thought *Possibly* Required for Successful Performance on (Implicit) Nonverbal TOM Test	T <Actor + Toy + Green Box>	1st

thought like the one in row 2 if she is to perform successfully. This thought has a second-order representational structure, since it is a thought about another mental state (the actor's belief).

In the third row we see a representation of a HOT that, if formed appropriately, would provide for basic state consciousness. Since this thought is also about another mental state (one's first order belief that all people are equal), this thought also has a second-order representational structure. And so again we see that the representational structure of the subject's thought about the actor's belief (as shown in row 2) is similar to the representational structure of the HOT that would afford basic state consciousness (as shown in row 3), and hence we see why one might be tempted to adopt the PR Assumption. If a thought like the one in row 2 is all that is required for successful

performance on these nonverbal TOM tests and if these thoughts have the same second-order representational structure as the thoughts required for basic state consciousness, then these tests would appear to measure precisely the capacity to represent mental states in thought that the HOT theory requires for basic state consciousness.

Now, it was at this stage in our assessment of the verbal TOM tests that we were forced to reject the PR Assumption, because we learned from Ruffman et al. (2001) that subjects needed to be *conscious* of what they thought about Sally's belief in order to pass those tests. In regards to these nonverbal TOM tests, however, the lesson from the Ruffman et al. study is quite different.

Recall, Ruffman et al. found that subjects could succeed on nonverbal TOM tests, by correctly looking to the location where someone will search for an object, without showing any evidence that they were *conscious* of the understanding of the mind that was expressed by their correct eye gaze behaviour. (Again, this was the lesson Ruffman et al. drew from the fact that the youngest split-responders failed to take into account the information expressed by their correct looking behaviour when betting on their incorrect verbal responses.) Since a subject's eye gaze was found to be measuring an *implicit* understanding of the mind in these cases, there would be no need for subjects in these nonverbal TOM tests to form further third-order thoughts in order to make their second-order representations of the actor's belief state conscious. So, in distinction from our conclusion about the standard verbal TOM tests, if a subject does need to represent the actor's belief in order to succeed on these nonverbal TOM tests, then that subject need only form a thought with a second-order representational structure (like the one

represented in row 2), in order to correctly anticipate the actor's searching and demonstrate this expectation with her eye gaze.

So far there are two interesting points to note here: First, it appears that, at this stage in our assessment at least, we are not yet forced to give up the PR Assumption. Since subjects in nonverbal TOM tests do not need to be conscious of the information tapped by the eye gaze measures in order to perform successfully on these tests, the formation of second-order thoughts about the actor's belief may be all that's required for successful performance. Since the second-order representational structure of these thoughts is parallel to the second-order representational structure of the HOTs required for basic state consciousness, the PR Assumption is not threatened.

Second, our earlier discussion of the HOT theory might actually provide an explanation of *why* successful performance is possible in these cases. Recall, we learned from the Awareness Principle that the formation of an unconscious mental state can enable a person to be differentially responsive to what that state represents, despite the fact that the person is not conscious of what she is representing. This was the accepted explanation of how blindsight patients perform successfully on forced-choice tests, for example, and this account seems to work just as well for explaining the performance of infants on these nonverbal TOM tests. Specifically, despite the fact that the infant is not conscious of what she thinks about the actor's belief in these cases, insofar as the infant does represent the actor's belief, the infant will be afforded an *awareness* of the actor's belief and so this unconscious representation will enable the infant to respond differentially on the basis of what she represents the actor to be believing.

Before celebrating any victories here, however, recall that our assessment so far is also based on a second assumption, namely, that an infant's successful performance in these nonverbal TOM tests *does* require, in the first place, that the infant represent the actor's belief about where the toy is located. As I mentioned when introducing this assumption, however, researchers are far from agreeing on this issue. In fact, there are numerous competing explanations of how infants achieve their success on nonverbal TOM tests, some of which appeal to mental state attribution but some of which do not require that an infant represent any mental states at all. If it turns out that infants can pass nonverbal TOM tests without representing any mental states at all, then we would no longer be justified in taking an infant's success on these tests as indicative of the infant's having the capacity to represent mental states in thought that the HOT theory requires. Hence the PR Assumption once again would have to be rejected.

Since we've just seen an account of how infants might perform successfully if they were attributing mental states to the actor, let's now consider an account of successful performance that does not require subjects to attribute mental states to the actor. Such an account is provided, for example, by Perner and Ruffman (2005).

Perner and Ruffman argue that infants can pass these nonverbal tests by applying a simple behavioural rule, such as 'agents tend to search for things in the last place they saw them located'. Perner and Ruffman argue that, though such a rule might only work because agents have minds and because certain mental states are causally connected with certain behaviours, it is possible for infants to formulate and make use of such rules, perhaps, for example, by initially extracting such rules from behavioural regularities,

without yet knowing anything about the mind or the causal connections between one's mind and one's behaviour. [30]

To explain, for example, how an infant might make use of a rule about 'seeing' without having any conception of mental states, Perner and Ruffman suggest that the infants instead simply might form "three-way actor-object-location associations" (Perner and Ruffman (2005), p. 215). They explain these associations as patterns of neuron firings (or, more simply, as representations) that encode united information about the actor, the search object, and the last location at which the actor had unobstructed eye contact with the object. So rather than requiring an infant to represent the actor's perceptual states or beliefs, the infant instead need only represent this connection between non-mental facts and objects. To put this in the terminology we have been using then, these sorts of representations would only have a *first-order* representational structure, since they are representations of non-mental facts and objects rather than representations of other mental states. An example of such a thought is presented in row 4 of Table 2.

Furthermore, once an infant forms such a three-way association, Perner and Ruffman argue that the representation can support the very same expectations of behaviour that were, on the competing account, taken to be supported by attributions of mental states. For example, if an infant forms a representation associating the actor, the toy, and the green box location, and does not form a representation associating the actor, the toy, and the yellow box location (perhaps because the actor's eyes were never directed toward the toy when it was located at the yellow box), then the infant has the

[30] For example, Perner and Ruffman write, "such a rule captures something implicit about the mind, because the rule only applies as a result of the mind mediating between seeing and acting. Nonetheless, infants can simply know the rule without any conception that the mind is the mediator" (Perner and Ruffman (2005), p. 215).

foundation for an expectation that the actor will search for the toy in the green box location.

Notice also that this expectation will then be able to generate the very same looking time results as were found in both of the nonverbal TOM experiments. In regard to the Onishi and Baillargeon (2005) study, since the infant expects the actor to search in the green box location (because the green box is represented in her three-way association), the infant would be surprised, and hence would look for longer, if the actor instead searched in the yellow box location. In regard to the Southgate et al. (2007) study, since the infant expects the actor to search in the green box location (because the green box is represented in her three-way association), the infant would likely look first and for longer overall at the green box location when the searching signal is sounded. Hence we see that the very same performance, which is based on the very same predictions of behaviour, can be grounded by these first-order three-way associations, without requiring that the infant represent any mental states at all. Furthermore, it's interesting to note that the authors of both studies actually mention this very rule and agree that their data could be sufficiently explained by an account such as this one.[31]

In both cases then, the very same data can be neatly explained by appeal to an infant's ability to track physical and behavioural data; we do not also need to suppose that an infant represents any of the actor's mental states. Furthermore, as we learned from the Ruffman et al (2001) study, infants would not need to be conscious of the information grounding their responses in these cases either, so it would not even be the case that they would have to make their own three-way-associations conscious by making them the

[31] Onishi and Baillargeon (2005), p. 257; Southgate et al. (2007), p. 591.

objects of higher-order thoughts. Hence in no way at all would this explanation require any sort of second-order representations.

In this way, we see how a purely behavioural rule (such as 'agents search for objects at the location where they last had unobstructed eye contact with that object') would account for the expectations and the varying looking times measured in these experiments, all the while not requiring that an infant attribute any sort of mental states to the actor or be able to represent mental states in thought at all. If this sort of explanation were accepted and we were to conclude that successful performance on nonverbal TOM tests does not require the formation of any second-order thoughts whatsoever, then the PR Assumption would have to be rejected. We would no longer be justified in taking success on nonverbal TOM tests as evidence that infants do have the very capacity to represent mental states in thought that the HOT theory requires for basic state consciousness and so the HOT theorists' own argument against the TOM objection would have to be rejected.

So, here we have two different ways of explaining the nonverbal TOM test data. According to one explanation infants do represent the mental states of others in order to make inferences about the other's future behaviour but, according to the other, infants need only represent non-mental facts about the relationship of people to objects and their environment in order make these inferences. Given that both explanations are just as successful at explaining the data, it would appear that the data themselves do not lend unique support to either interpretation. Hence the data cannot help us in determining whether the PR Assumption is warranted in these cases.

In light of this fact it would appear that the most conservative conclusion to draw would be to say that the PR Assumption is not clearly true and thus that we are not warranted in adopting it at this time. Since we are not safe in assuming that successful performance on nonverbal TOM test requires a subject to form a second-order thought, we cannot take these tests as indications of whether or not a creature has the metarepresentational capacities the HOT theory requires for basic state consciousness. Thus we come to the, perhaps tentative, conclusion that a subject's success on non-verbal TOM tests provides no clear evidence in support of the HOT theory.[32]

5. Conclusion

I have argued that both sides of this debate rely on the same false assumption, namely, that TOM tests require for success the same capacities the HOT theory requires for basic state consciousness. To support this conclusion I first demonstrated that this PR Assumption was necessary if either side was to argue that evidence from TOM tests has any bearing on the HOT theory in the first place. I then demonstrated that, in relation to the verbal TOM tests on which the TOM objection is based, the PR Assumption must be false because these tests require subjects to form third-order thoughts in order to be successful. If this analysis is sound then objectors cannot appeal to the *failure* of subjects on these verbal TOM tests as indications that the subjects lack what the HOT theory requires for basic state consciousness. Hence the TOM objection was defeated.

---

[32] Interestingly, even if we did assume that these nonverbal tests really do track the metarepresentational capacities the HOT theory requires for basic state consciousness, it turns out that there are still groups of people who fail even these non-verbal TOM tests yet, presumably, these people still have conscious mental states. For example, in a study identical to the one run by Southgate et al. (2007), Senju et al. (2010) found that children with Autism Spectrum Disorders (ASDs) failed to look in anticipation at the locations predicted by attributing false beliefs. Instead their performance was no different from chance. As the researchers write, "…children with ASD…fail to spontaneously anticipate others' actions when such anticipation requires the attribution of a false belief to the actor" (Senju et al., 2010, p. 359). Thus, even if we take the PR Assumption to be justified relative to these tests, the TOM objection still might not be defeated.

I then demonstrated that, in relation to the nonverbal TOM tests on which the HOT theorists' own argument against the TOM objection is based, the PR Assumption must be rejected, at least for now, be rejected because it is currently unclear whether these tests require subjects to form second-order thoughts or merely first-order thoughts in order to be successful. If this analysis is sound then proponents of the HOT theory cannot appeal to the *success* of subjects on these nonverbal TOM tests as indications that the subjects do have the metarepresentational capacities the HOT theory requires for basic state consciousness. Hence the HOT theorists' own argument against the TOM objection must also be rejected.

In sum, current TOM tests cannot be taken as measures of the capacity to represent mental states in thought that the HOT theory requires for basic state consciousness. The PR Assumption therefore must be rejected and so the arguments discussed herein, the TOM Objection and the HOT theorists' argument against the TOM objection, both must be rejected.

References

Baron-Cohen, S. (1995) *Mindblindness: An Essay on Autism and Theory of Mind*, Cambridge, MA: MIT Press.

Block, N. (1995) "On a Confusion About the Function of Consciousness", *Behavioral and Brain Sciences, 18(2)*, p. 227-247.

Block, N. (2011) "The Higher Order Approach to Consciousness Is Defunct", *Analysis, 71(3)*, p. 419-431.

Carruthers, P. (1989) "Brute Experience", *Journal* o*f Philosophy*, *86*, p. 258-269.

Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge: Cambridge University Press.

Carruthers, P. (2009) "How We Know Our Own Minds: The Relationships Between Mindreading and Metacognition", *Behavioural and Brain Sciences*, *32*, p. 121-182.

Chalmers, D. (1995) "Facing Up to the Problem of Consciousness", *Journal of Consciousness Studies*, *2(3)*, p. 200-219.

Clements, W. A. and Perner, J. (1994) "Implicit Understanding of Belief", *Cognitive Development*, *9*, p. 377-395.

Dretske, F. (1995) *Naturalizing the Mind*, Cambridge, MA: MIT Press.

Fodor, J. A. (1992) "Discussion: A Theory of the Child's Theory of Mind", *Cognition*, *44*, p. 283-296.

Frith, U. (2003) *Autism: Explaining the Enigma*, second edition, Malden, MA: Blackwell Publishing.

Gennaro, R. J. (1993) "Brute Experience and the Higher-Order Thought Theory of Consciousness", *Philosophical Papers, 22*, p. 51-69.

Gennaro, R. J. (1996) *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*, Amsterdam: John Benjamins Publishing Company.

Gennaro, R. J. (2004) "Higher-Order Thoughts, Animal Consciousness, and Misrepresentation", In R. J. Gennaro (ed.), *Higher-Order Theories of Consciousness: An Anthology*, Philadelphia, PA: John Benjamins North America, p. 45-66.

Happé, F. G. E. (1995), "The Role of Age and Verbal Ability in the Theory of Mind Task Performance of Subjects with Autism", *Child Development 66(3)*, p. 843-855.

Hare, B., Call, J., and Tomasello, M. (2001) "Do Chimpanzees Know What Conspecifics Know and Do Not Know?", *Animal Behavior*, *61*, p. 139-151.

Lurz, R. W. (2006) "Conscious Beliefs and Desires: A Same-Order Approach" In Uriah Kriegel and Kenneth Williford (Eds.) *Self-Representational Approaches to Consciousness*, Cambridge, MA: MIT Press, p. 321-351.

Lycan, W. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.

Nagel, T. (1974) "What Is It Like to Be a Bat?", *Philosophical Review, 83(4*), p. 435-450.

Onishi, K. H. and Baillargeon, R. (2005) "Do 15-Month-Old Infants Understand False Beliefs?", *Science, 308*, p. 255-258.

Perner, J. and Ruffman, T. (2005) "Infants' Insight into the Mind: How Deep?", *Science, 308*, p. 214-216.

Perner, J., Leekam, S. R., and Wimmer, H. (1987) "Three-Year-Olds' Difficulty with False Belief: The Case for a Conceptual Deficit" *British Journal of Developmental Psychology, 5*, 125-137.

Povinelli, D.J., Nelson, K.E., and Boysen, S.T. (1990) "Inferences about Guessing and Knowing by Chimpanzees (*Pan troglodytes*)", *Journal of Comparative Psychology*, *104*, p. 203-210.

Premack, D. and Woodruff, G. (1978) "Does the Chimpanzee Have a Theory of Mind?" *The Behavioral and Brain Sciences*, *4*, 515-526.

Rosenthal, D. M. (2000) "Consciousness and Metacognition", In Dan Sperber (Ed.) *Metarepresentations: A Multidisciplinary Perspective*, New York, NY: Oxford University Press, p. 265-295.

Rosenthal, D. M. (2002a) "A Theory of Consciousness", In Ned Block, Owen Flanagan, and Güven Güzeldere (Eds.) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press, p. 729-753.

Rosenthal, D. M. (2002b) "Explaining Consciousness", In David J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford: Oxford University Press, p. 406-421.

Rosenthal, D. M. (2003) "Unity of Consciousness and the Self", *Proceedings of the Aristotelian Society*, *103*, p. 325-352.

Rosenthal, D. M. (2004) "Varieties of Higher-Order Theory", In Rocco J. Gennaro (Ed.) *Higher-Order Theories of Consciousness: An Anthology*, Amsterdam: John Benjamins Publishing Company, p. 17-44.

Rosenthal, D. M. (2005) *Consciousness and Mind*, Oxford: Clarendon Press.

Rosenthal, D. M. (2011) "Exaggerated Reports: Reply to Block" *Analysis, 71(3)*, p. 431-437.

Ruffman, T., Garnham, W., Import, A., and Connolly, D. (2001) "Does Eye Gaze Indicate Implicit Knowledge of False Belief? Charting Transitions in Knowledge" *Journal of Experimental Child Psychology*, *80*, p. 201-224.

Seager, W. (2004) "A Cold Look at HOT Theory", In R. J. Gennaro (Ed.), *Higher-Order Theories of Consciousness: An Anthology*, Philadelphia, PA: John Benjamins North America, p. 255-275.

Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., Osanai, H., and Csibra, G. (2010) "Absence of Spontaneous Actions Anticipation by False Belief Attribution in Children with Autism Spectrum Disorder" *Developmental Psychology*, *22*, p. 353-360.

Southgate, V., Senju, A., and Csibra, G. (2007) "Action Anticipation Through Attribution of False Belief by 2-Year-Olds", *Psychological Science*, *18(7)*, p. 587-592.

Weisberg, J. (2011) "Abusing the Notion of What-It's-Like-Ness: A Response to Block" *Analysis, 71(3)*, p. 438-443.

Weiskrantz, L., Warrington, E. K., Sanders, M. D., and Marshall, J. (1974) "Visual Capacity in the Hemianopic Field Following a Restricted Occipital Ablation" *Brain*, *97*, p. 709-728.

Wellman, H. M. (1991) "From Desires to Beliefs: Acquisition of a Theory of Mind", In Andrew Whiten (Ed.) *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading*, Cambridge, MA: Basil Blackwell Inc., p. 19-38.

Wimmer, H. and Perner, J. (1983) "Beliefs about Beliefs: Representations and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception" *Cognition*, *13*, p. 103-128.

Paper 2: ***Why the Higher-Order Thought Theory*** **Cannot** ***Claim that Basic State Consciousness Involves Higher-Order Phenomenal Character***

1. Introduction

Robert Lurz (2003, 2006) convincingly argues that the phenomenal character affiliated with basic state consciousness does not necessarily involve one's becoming higher-order conscious *of the fact that one has a particular mental state*. I will call this Lurz's 'Phenomenal Character Argument'. Beyond merely establishing this fact however, Lurz believes his Phenomenal Character Argument also provides good reason to reject the Higher-Order Thought (HOT) Theory of Consciousness. Since Lurz takes HOT theorists to claim that the phenomenal character affiliated with basic state consciousness *does* necessarily involve one's becoming higher-order conscious of the fact that one has a particular state, and since his Phenomenal Character Argument shows this not to be the case, Lurz takes his argument to show that the HOT theory fails to accurately account for phenomenal character and hence that the theory can be rejected.

I will argue that Lurz's interpretation of the account of phenomenal character provided by the HOT theory is incorrect and hence that his argument against the HOT theory fails. In fact, I argue that the HOT theory *cannot* provide the characterization of phenomenal character that Lurz interprets it to be providing, since the HOT theory *cannot* say that the phenomenal character affiliated with *basic* state consciousness involves one's becoming conscious of what one's (unconscious) HOTs represent.

I also have a few peripheral goals in this paper. In presenting the argument outlined above, I hope to draw attention to a fact about the HOT theory that appears to go

unnoticed, namely, that there is an important distinction between the HOT theory's account of *what grounds* state consciousness and the HOT theory's account of *what it's like to have* state conscious states. I contend that only the former account can be provided in terms of higher-order representation, so the HOT theory's two accounts must be importantly distinct.

Finally, I hope that this paper will initiate a serious philosophical conversation about the HOT theory's account of phenomenal character. There is currently very limited discussion of this aspect of the HOT theory and, though I believe the interpretation I offer here is sound, I present this paper as a call to those in support of, and those against, the HOT theory to join in this discussion.

The paper will proceed as follows: Since one aim of this paper is to spark a conversation about the HOT theory's account of phenomenal character, I begin by identifying some of my assumptions and by clarifying some of the terminology with which the issues in this paper will be discussed. The aim is to make clear any details that may bring confusion to the debate. I then briefly outline the relevant aspects of the HOT theory, focusing solely on David Rosenthal's (2002a, 2002b, 2005) version of the theory since Lurz himself specifically targets Rosenthal's account. Next I present Lurz's interpretation of the HOT theory's account of phenomenal character and I present Lurz's Phenomenal Character Argument against the HOT theory. I then explain why HOT theorists *cannot* be explaining the phenomenal character affiliated with basic state consciousness in the way Lurz understands them to be explaining it and hence I demonstrate that Lurz's interpretation must be incorrect. I conclude that we can therefore reject Lurz's Phenomenal Character Argument against the HOT theory. Finally, I discuss

the implications of my argument for the separation of the HOT theory's accounts of what state consciousness consists in and of what it's like to have state conscious states, and I briefly sketch an alternative interpretation of the HOT theory's account of phenomenal character, providing textual support for this new account.

2. Setting the Terms of the Debate

Since I hope this paper will initiate a philosophical discussion of the HOT theory's account of phenomenal character, and since there are many issues involved in such a discussion that are quite complicated in their own right, I'd like to begin this conversation by clearly outlining some of the assumptions I make in this paper and by clarifying some of the terminology I will use throughout the discussion. In starting the conversation in this way, I hope to engage fellow philosophers with minimal confusion.

First of all, for the purposes of this paper we will assume that when we say someone has a *conscious* mental state, we are saying that the person has a mental state that instantiates the property of *state consciousness*. Furthermore, we will understand this property of state consciousness to be a relational property consisting in the intentional relationship between a first-order state and an appropriate higher-order thought that represents that first-order state.[1] Since this is the characterization of state consciousness that HOT theorists most commonly present[2], and since this is the characterization that Lurz also seems to work with, it will suffice for our purposes here.

---

[1] The precise meaning of this claim will become clearer once I introduce the HOT theory in the next section.

[2] Though HOT theorists do work with this characterization of state consciousness most often (for example, see Rosenthal (2002a)), they also suggest that this characterization might not be entirely accurate. (For example, see Rosenthal (2000, 2003, and 2011).) Since speaking in terms of the more accurate characterization will unnecessarily complicate our discussion here, however, and since HOT theorists themselves often avoid the more accurate characterization for the very same reason, I feel comfortable doing so here as well. I do believe that my arguments in this paper are compatible with the more strict account of state consciousness, though I do not intend to provide any argument to that effect here.

Second, there is agreement among those in the debate that in saying that a mental state is state conscious, we mean to identify that there's *something it's like for* the person who instantiates that state. The phrase 'something it's like for…' was introduced by Thomas Nagel (1974) and has since become a common way of loosely and intuitively describing the essential subjective and experiential aspect of consciousness, for example, the *redness* one experiences when one consciously sees a red tomato or the *painfulness* one feels when one consciously pinches a finger in a door. This essential, subjective nature of consciousness is also sometimes referred to as the *qualitative* or *phenomenal* character of consciousness or simply as *what-it's-like-ness*. I will use these terms interchangeably throughout our discussion and I ask the reader to note that I intend no theoretical ties to any one particular theory of phenomenal character when I use terms like 'phenomenal character' or any of the others mentioned here.[3]

Third, we will follow Lurz (2006) in assuming that there will be some sort of phenomenal character affiliated with a creature's having *any* sort of state conscious mental state. So, there will be phenomenal character affiliated with one's having conscious sensations and perceptions as well as with one's having conscious propositional attitudes, like conscious beliefs or conscious desires.

The reader should note, however, that it is unclear whether or not Rosenthal would agree with this assumption. On the one hand, he often says that the phenomenal character affiliated with state conscious sensations and perceptions is due to one's representing, via HOTs, various special, qualitative properties that are instantiated by

[3] For example, despite mentioning Block's (2011) term for this aspect of experience (i.e., 'what-it's-like-ness'), I do not intend to have any ties to the specific theory of phenomenal consciousness that Block himself endorses (see Block 1995). Instead, I just mean to identify what Weisberg (2011) calls the "moderate" reading of terms like 'phenomenal consciousness' and 'what-it's-like-ness'.

only sensory and perceptual states.[4] It might appear, then, that Rosenthal would *not* endorse the claim that those states that are presumed to lack these qualitative properties (in particular, propositional attitude states) would have any phenomenal character affiliated with them when they are state conscious.

On the other hand, Rosenthal also says, for example, that, "[a] state's being conscious is a matter of mental appearance – of how one's mental life appears to one. …a state is conscious only if one is subjectively aware of oneself as being in that state" (Rosenthal (2011), p. 431). With this sort of description of phenomenal character, a description which makes no appeal to qualitative properties *per se* but rather only to one's subjective awareness of mental states, there would seem to be no reason to deny that one has a subjective appearance affiliated with being in a conscious belief in addition to a subjective appearance affiliated with being in a conscious sensory state. Since Rosenthal's position on the issue is unclear, and since our aim in this paper is to engage with Lurz, who does assume there will be some sort of phenomenal character affiliated with a creature's having *any* sort of state conscious mental state, we therefore will adopt the assumption that there will indeed be something it's like for a creature who instantiates any kind of state conscious mental state.

So our assumptions are as follows: When we say a mental state is state conscious we mean to identify not only that there is a relational property that we attribute to that mental state, namely, state consciousness, but also that there is something it's like, an affiliated phenomenal character, for the bearer of that state. We can summarize these

[4] See, for example, Rosenthal (2004).

points neutrally by saying that *attributions of state consciousness must entail attributions of what-it's-like-ness*.

Finally, as mentioned in the introduction, I believe there are two separate aspects to the HOT theory's account of consciousness and that these aspects are rarely distinguished from one another. I would like to take a moment to explain the distinction between these aspects now.

We can differentiate these two aspects by means of the following two questions: (1) *What is it like for a creature* who has a state conscious mental state?; and (2) *What grounds* state consciousness?. Let's briefly discuss how these two questions differ.

Someone who poses the first question, '*What is it like for a creature who has a state conscious mental state?*', is asking for a description of phenomenal character. For example, one might ask, what's it like for you to consciously be in pain right now, and you might answer that you have a dull ache, or a throbbing pain, or a searing pain, and so on. By characterizing your pain as dull, throbbing, or searing, you are identifying aspects of phenomenal character – aspects of what it's like for you as you have this conscious pain.

On the other hand, someone who poses the second question, '*What grounds state consciousness?*', is not asking for an account of your phenomenal character. Rather, the person is asking for an account of what makes it the case that your mental state is state conscious rather than unconscious. Again, for example, they might be asking what makes your pain count as being a conscious pain rather than an unconscious pain and you might answer by explaining what it is that generates, realizes, or subvenes your pain's property of state consciousness. To my mind, this is an importantly different question

than the first and, as we'll see, I think the HOT theory provides importantly distinct answers to each of these questions.

I note these things here because the distinctions can be subtle and the phrases required to describe these different accounts can get complicated. So, to be clear, when I speak of characterizing the phenomenal character affiliated with state consciousness or of what it's like for someone who has a state conscious state, I intend to be discussing facts related to the first of these two questions. On the other hand, when I speak of what grounds state consciousness or of what state consciousness consists in, I intend to be discussing facts related to the second of these two questions.

With these assumptions now explicit and our terminology clarified, we are ready to explore the HOT theory's account of state consciousness.

3. The Higher-Order Thought Theory on the *Constitution* of State Consciousness

The HOT theory is most often presented as an account of what *constitutes* state consciousness. For example Rosenthal sets up a paper introducing the HOT theory by writing,

> Assuming that not all mental states are conscious, we want to know how the conscious ones differ from those which are not. And, even if all mental states are conscious, we can still ask what their being conscious *consists* in. We can call this the question of state consciousness. This is my main concern in what follows.
>
> (Rosenthal (2002a), p. 729, emphasis added.)

In order to determine what state consciousness consists in, HOT theorists begin with a simple insight from the everyday folk: We find it natural to say that if a person is not aware of her mental state in any way, then her mental state is not conscious. This implies, conversely, that a conscious mental state must be a state a person is aware of in some suitable way. HOT theorists take this folk intuition to entail that the difference

between a mental state when it's state conscious and when it's unconscious lies in its possessor's awareness of it, and so they set out to explain the property of state consciousness in terms of a person's awareness of her own mental states. HOT theorists hypothesize that a person comes to have this sort of awareness by forming higher-order thoughts about her mental states and so they argue, specifically, that a mental state is state conscious only when that mental state is represented by an appropriate higher-order thought (HOT).[5] Let's unpack this claim a little further.

First, note that the higher-order thoughts required for state consciousness are called *higher-order* thoughts simply because they are mental representations of other mental representations. It is common practice, for example, to refer to representations of non-mental objects or facts as *first-order* representations, to refer to representations of first-order states as *second-order* representations, to refer to representations of second-order states as *third-order* representations, and so on. Any mental state that is not a first-order representation counts as being a higher-order representation.

Second, we saw above that not just any higher-order thought will suffice for state consciousness, the HOT has to be *appropriate*. According to Rosenthal, a HOT is *appropriate* when it is noninferential (i.e., it is not the product of any conscious inference or observation), when it is nondispositional (i.e., when a person actually instantiates the HOT rather than merely having a disposition to form such a HOT), when it is assertoric (i.e., when the propositional attitude of the thought is one of assertion), and when the

[5] For example, Rosenthal writes, "On higher-order theories, first-order states do not inherit the property of being conscious from higher-order states. On such theories, the property of a state's being conscious *consists of* one's being aware of oneself as being in that state, and the higher-order states *constitute* those awarenesses. The HOA [i.e. Higher-Order Awareness] does not pass along the property of being conscious to the first-order state; it simply serves to make one aware of that state in the right way, and that is what the state's being conscious *consists* of" (Rosenthal (2012), p. 1428, emphasis added.)

HOT represents its bearer as being in a particular mental state (for example, only an appropriately formed HOT with the content, roughly, "I believe that there's an apple on the table" will result in the conscious belief that there is an apple on the table). Though all of these conditions are important components of the theory, only one will be particularly relevant to our discussion here – the noninferential condition – so let's look at that one in a little more detail.[6]

One common way of characterizing the phenomenal character affiliated with state consciousness is by saying that some content or information comes to be suddenly and immediately before one's mind. It's taken to be a fact up for explanation that the phenomenology appears sudden and immediate, rather than being mediated in some way, and the HOT theory explains this phenomenological fact by positing that the HOTs that are involved in making our mental states state conscious must not be arrived at by means of conscious inference or observation. If we came to have our HOTs via a process of conscious inference or observation (for example, perhaps we recollect our own recent behaviour and surmise that we must be anxious about an impending deadline), then insofar as we are conscious of the inferences or observations that mediated the process of coming to be aware of our mental states, the resulting awareness we have of our mental states would no longer seem to be sudden and immediate.

This fact about phenomenological immediacy also points to another important aspect of the HOT theory, the claim that the HOTs which constitute the requisite awareness of one's mental state will, themselves, not be state conscious mental states. This claim helps us account for the phenomenological immediacy of conscious

[6] For a concise discussion of the conditions that make a HOT appropriate as well as the arguments leading to the account of state consciousness presented here see Rosenthal (2002b), especially Section II, "The Hypothesis", pg. 408-411.

experience because, again, it's a fact up for explanation that it generally never seems to us that there is a HOT mediating our awareness of our state conscious mental states. Since the HOT theory does say that HOTs mediate this process, they therefore also must claim that there is not anything it's like for someone to be instantiating those HOTs, else the person would notice those HOTs as mediators. Hence HOT theorists argue that the HOTs required for state consciousness are not themselves state conscious mental states.

This is not to say that HOTs can never be state conscious mental states however. In fact, whether or not one's HOTs are state conscious will be the determining factor between whether a state counts as being state conscious in what we'll call the *basic* sense, or whether that state counts as being state conscious in the *introspective* sense.

According to HOT theorists, *basic* state consciousness occurs when a first-order mental state (i.e., a mental state that represents facts about or objects in the world) is represented by an appropriate second-order thought (a thought about that first-order state). In this case, the second-order thought itself will be an unconscious state because there is no further HOT formed about it. On the other hand, *introspective* state consciousness occurs when this second-order mental state (i.e., the HOT involved in basic state consciousness) is represented by a third-order thought (a thought about the second-order thought about the first-order state). Again, this third-order thought itself will not be state conscious in these cases, because it is not represented by a higher-order thought.[7]

So, for example, if George believes that there is an apple on the table, and if his belief is to be state conscious in the basic sense, HOT theorists would say George must

[7] For a concise description of these accounts of basic and introspective state consciousness and the differences between the two see, for example, Rosenthal (2005), p. 48-49.

also form an appropriate second-order thought with the content, roughly, "I believe that there is an apple on the table". If George's belief is to be introspectively state conscious, however, he would need to form not only the first-order belief and a second-order thought about himself as having that belief, but also a third-order thought about the second-order thought, so a thought with the content, roughly, "I think that I believe that there is an apple on the table".

And so we see the HOT theorists' account of what constitutes state consciousness. A mental state, *M*, is state conscious if and only if *M* is represented by an appropriate higher-order thought. *Basic* state consciousness involves the formation of a second-order thought that itself is not state conscious, and *introspective* state consciousness involves the formation of a third-order thought that renders the second-order thought state conscious.

As mentioned, the HOT theory also provides an account of the phenomenal character affiliated with state consciousness. As we will need to understand Lurz's interpretation of this account in order to understand his Phenomenal Character Argument against the HOT theory, let's turn to his interpretation now. (I will present my own alternative interpretation at the end of the paper.)

4. Lurz's Interpretation: The Higher-Order Thought Theory on the *Phenomenal Character* of State Consciousness

As noted earlier, when we say that *there's something it's like for a creature*, we are saying that there's some content or information that seems suddenly and immediately to present itself before one's mind. Hence, in order to characterize what it's like for a creature, we would have to characterize what the creature suddenly and immediately

becomes conscious *of* in such instances. It is therefore in these terms that we will discuss the HOT theory's account of phenomenal character.

Because the HOT theory postulates that state consciousness *consists* in the awareness we have of our mental states (as a result of forming appropriate HOTs about those states), Lurz believes HOT theorists also explain the phenomenal character affiliated with state consciousness in terms of this higher-order awareness. Specifically, Lurz takes the HOT theorist to be claiming that what it's like for someone who has a state conscious mental state, *M*, can be captured by saying that the bearer of *M* suddenly and immediately becomes conscious *of what her HOT represents*, namely, *of the fact that she has M*.

Now, there's no question why Lurz might interpret HOT theorists to be giving such an account. Rosenthal does, after all, appear to give precisely this sort of account as he often says things like, "…a mental state's being conscious consists in our being conscious of being in that state" (Rosenthal, 2002a, p. 745). In light of explicit statements like these, Lurz does seem justified in taking HOT theorists to be explaining the phenomenal character affiliated with state consciousness in terms of one's becoming conscious of the fact that one has a particular lower-order state.

Since Lurz's argument is specifically targeted at the HOT theory's account of the phenomenal character affiliated with *basic* state consciousness, let's spell out an example of this account in those terms. We saw that, according to the HOT theory, for George's first-order belief that there's an apple on the table to be state conscious in the basic sense, George must form a second-order thought with the content, roughly, "I believe that there's an apple on the table". Lurz interprets the HOT theory also to be saying that

when George has this state conscious first-order belief, what it's like for George is that he suddenly and immediately becomes higher-order conscious *of the fact that he believes that there is an apple on the table*, so he becomes conscious of the very fact represented by his second-order thought. By claiming that the consciousness in this case is *higher-order*, Lurz simply means that what George becomes conscious *of* is a fact about his own mental states, namely, the fact that he has a particular belief that there is an apple on the table.

We can contrast this claim, for example, with the claim that the phenomenal character affiliated with basic state consciousness is instead characterized by saying that one becomes *first-order* conscious of objects in or facts about the world. According to Lurz (2003), Fred Dretske (1995) provides such an account. The consciousness here is *first-order* because it is a consciousness of facts or objects in the world, rather than a consciousness of mental facts or objects. On this account then, when George's first-order belief is state conscious in the basic sense, what it's like for George is that he suddenly and immediately becomes first-order conscious *of the apple on the tabl*e or, perhaps, *of the fact that there is an apple on the table*.

Both of these accounts can be contrasted with the claim that the phenomenal character affiliated with basic state consciousness is characterized by one's becoming *same-order* conscious of what one's mental states represent. Lurz (2003, 2006) himself promotes such a view. Lurz refers to this sort of consciousness as being *same-order* because one is said to become suddenly and immediately conscious of precisely the *same* representational content that is expressed by one's state conscious mental state itself, as opposed to becoming first-order conscious of the facts represented by that state conscious

state and as opposed to becoming higher-order conscious of that fact that one has that state conscious state.[8] On this account then, when George's first-order belief is state conscious in the basic sense, what it's like for George is that he becomes same-order conscious *of* the propositional content expressed by his belief, namely, the proposition *that there is an apple on the table*.

So, according to Lurz each of these accounts makes a different claim about the phenomenal character one must experience when one has a state conscious state in the basic sense. These differences in phenomenal character are captured by the differences in what one is said to become conscious *of* when one has a state conscious mental state. Importantly, as Lurz understands the HOT theory, it claims that one must become conscious *of the fact that one has a particular first-order state* – the very fact represented by one's second-order thought – when one comes to have a mental state that is state conscious in the basic sense.[9] With this in mind, we are ready to see where Lurz thinks the HOT theory goes wrong with this account of phenomenal character.

5. Lurz's *Phenomenal Character Argument* Against the Higher-Order Thought Theory of Consciousness

The essence of Lurz's Phenomenal Character Argument is as follows: Lurz argues that the HOT theorists' account of the phenomenal character affiliated with basic state consciousness is incorrect because someone can have a mental state that is state conscious in the basic sense without being higher-order conscious of the fact that she has

[8] Specifically, Lurz argues that we become conscious of the representational content of our mental states by forming an appropriate deictic demonstrative belief that identifies the representational content of those mental states. For example, in order for George's belief to be state conscious, George must form a separate belief with the content "It's that there is an apple on the table", where the 'it's' here demonstratively refers to the proposition expressed by George's first-order belief and the rest identifies the actual proposition expressed by George's first-order belief about the apple. For more on the details of Lurz's account see Lurz (2003, 2006).

[9] For further discussion of his interpretation of the HOT theory see, for example, Lurz (2006), p. 325-328.

that mental state. This means that HOT theorists were wrong in saying that the phenomenal character affiliated with basic state consciousness *must* involve one's becoming higher-order conscious of what one's HOT represents, and hence this demonstrates that the HOT theory is mistaken. Since this aspect of the account fails, Lurz concludes that the HOT theory should be rejected.

Lurz's argument proceeds in roughly two stages: First he shows that higher-order consciousness of our own mental states is not a necessary component of the phenomenal character affiliated with state consciousness. Then he concludes, given his interpretation of the HOT theory's account of phenomenal character, that the HOT theory must be mistaken and can therefore be rejected.

To support the first half of this argument, Lurz (2006) presents an intuitive example of someone's being in a mental state that is state conscious in the basic sense while she also fails to be higher-order consciousness of the fact that she has that very state. This example helps Lurz establish his conclusion that a higher-order consciousness of what one's HOT represents cannot be a necessary component of the phenomenal character affiliated with basic state consciousness.

Lurz's (2006) example is as follows: Imagine a colleague informs you that someone has stolen your backpack out of your office. Perhaps you're not initially all that upset because the backpack wasn't that valuable, you had removed the books from it, and so on. After a moment, however, it dawns on you – your keys are in that bag! You then rush to call the campus police to try to get your backpack returned.

Lurz thinks there are two important facts established by this example. First, at the moment, call it *t*, that it dawns on you that your keys are in your bag, it's clear that there

comes to be something it's like for you to believe that your keys are in your bag, hence we should agree that your belief that your keys are in your bag becomes state conscious at *t*. You might have formed that belief earlier in the morning, say, when you put your keys in the bag, but that belief was not state conscious at the moment immediately prior to time *t*; instead, it came to be state conscious (or maybe, came to be state conscious again) only at time *t*.

The question is then how to properly characterize the phenomenal character affiliated with this state conscious belief. Lurz's answer is as follows:

> …what seems to be going on is that at time *t* I suddenly come to be actually immediately aware *of* something, and this something seems to be *what* I believe…not, that I *believe*…something. …I seem to become at time *t* actually immediately aware *of what* I believe with respect to the whereabouts of my keys, for at time *t*, what I believe with respect to the whereabouts of my keys – namely, that my keys are in my bag – suddenly occurs to me in a way which, from my point of view, did not involve any inference or observation on my part. Now, if what one believes when one believes that *p*…is the proposition that *p*…, then what…I seem to become actually immediately aware of is a proposition. …it is the proposition *that my keys are in my bag*…
>
> (Lurz (2006), p. 332, original emphasis.)[10]

Here we see the second fact established by this example. Lurz argues that what he is conscious *of* at time *t* is actually *what* he believes about the location of his keys, i.e., the proposition 'that my keys are in my bag'.[11] Importantly, it does seem that this

---

[10] At the end of the paper I introduce a distinction between *consciousness* and *awareness*. The astute reader might notice that in the passage quoted here Lurz uses the term 'aware' rather than the term 'conscious'. Though he does not say this explicitly, and hence I am noting the issue for the reader, Lurz (2006) appears to use these two terms interchangeably and so he does not appear to be sensitive to the sort of distinction I will introduce later on.

[11] Lurz does provide an argument for his claim that we become conscious of the *proposition* expressed by our conscious beliefs, rather than becoming conscious of the worldly facts represented by our beliefs. Since this argument is not central to our discussion however, I will refrain from presenting it here. I refer the interested reader to Lurz (2003, 2006) for further details of that argument.

characterization captures the phenomenal character that would be affiliated with such belief becoming state conscious in this scenario. This seems clear when you imagine yourself in the example. There's a sudden and overwhelming sinking sensation – my keys are in that bag! – and as you imagine experiencing that realization, and imagine the experience of it sinking in, notice that it seems true that you would not also have an explicit higher-order consciousness of the fact that you are *believing* that your keys are in the bag. Hence Lurz's example does seem to characterize the phenomenal character affiliated with this conscious belief, and it does seem to do so without making any mention of one's becoming conscious *of the fact that one believes that one's keys are in one's bag*.

Furthermore, since we have already established that the belief in this scenario counts as a state conscious belief at moment *t*, we also can see now that this example demonstrates precisely what the HOT theory, as interpreted by Lurz, should predict *not* to be possible – one comes to be same-order conscious of what one's state conscious belief represents (i.e., the proposition 'that my keys are in my bag') without also becoming higher-order conscious of the fact that one believes that one's keys are in one's bag. Since, at *t*, one would both have a state conscious belief (in the basic sense) and yet fail to be higher-order consciousness of the fact that one has that belief, so it would appear that being higher-order conscious of what one's HOT represents cannot be a necessary component of the phenomenal character affiliated with basic state consciousness after all.

The last step in the argument then follows simply: Since Lurz takes the HOT theory to be saying that this sort of higher-order consciousness is a necessary component of the phenomenal character affiliated with basic state consciousness, and since this

example shows that it is not, Lurz concludes that the HOT theory must provide an incorrect account of phenomenal character. We can summarize Lurz's Phenomenal Character Argument against the HOT theory as follows:

(1) If we can characterize the phenomenal character affiliated with a person's having a state conscious mental state, in the basic sense, without saying that she becomes higher-order conscious of the fact that she has that mental state, then becoming higher-order conscious of the fact that one has a particular mental state cannot be a necessary component of the phenomenal character affiliated with basic state consciousness.

(2) We can characterize the phenomenal character affiliated with a person's having a state conscious mental state, in the basic sense, without saying that she becomes higher-order conscious of the fact that she has that mental state.

This point is demonstrated by the stolen backpack example. Following from (1) and (2) we get:

(3) Therefore, becoming higher-order conscious of the fact that one has a particular mental state cannot be a necessary component of the phenomenal character affiliated with basic state consciousness.

(4) The HOT theory says that one must form an appropriate HOT about oneself as being in a first-order state for that first-order state to be state conscious in the basic sense.

This is the HOT theorists' explanation of what constitutes state consciousness.

(5) If the HOT theory says that one must form an appropriate HOT about oneself as being in a first-order state for that first-order state to be state conscious in the basic sense, then the HOT theory also must be claiming that higher-order consciousness of the fact that one has a particular mental state is a necessary component of the phenomenal character affiliated with basic state consciousness.

I take it that this is a charitable way of connecting the claims actually made by HOT theorists with Lurz's particular interpretation of the theory's account of phenomenal character. Following from (4) and (5) we get:

(6) Therefore, the HOT theory must be claiming that higher-order consciousness of the fact that one has a particular mental state is a necessary component of the phenomenal character affiliated with basic state consciousness.

(7) If the HOT theory claims that higher-order consciousness of the fact that one has a particular mental state is a necessary component of the phenomenal character affiliated with basic state consciousness and yet this claim is incorrect, then the HOT theory must be mistaken and can be rejected.

(8) Therefore, the HOT theory must be mistaken and can be rejected.

This final conclusion follows from (3), (6), and (7). With the logic of Lurz's argument now clear, we are ready to see why the argument fails. Let's move to that analysis now.

6. An Analysis of Lurz's *Phenomenal Character Argument*

In the first half of his argument, Lurz demonstrates that a person's being higher-order conscious of the fact that she has a particular lower-order state is not a necessary component of the phenomenal character affiliated with basic state consciousness. Instead, it would appear that a characterization in terms of one's same-order consciousness of what one's mental state represents is a sufficient characterization of the phenomenal character affiliated with basic state consciousness. I am willing to grant Lurz these conclusions.

In the second half of his argument, Lurz argues that we are forced to reject the HOT theory because it endorses an incorrect account of the phenomenal character affiliated with basic state consciousness. This is where I disagree. Lurz's argument for this further conclusion rests on an assumption (represented as premise 5 in my reconstruction of his argument above) that, in saying one must form a HOT for one to have a mental state that is state conscious in the basic sense, HOT theorists also must be claiming that we cannot characterize what it's like for someone who has a state conscious

state in the basics sense, without saying that she becomes higher-order conscious of the fact that she has a particular first-order state. I think we can, and should, reject this assumption. In fact, I will argue that HOT theorists *cannot* claim that the phenomenal character affiliated with basic state consciousness involves one's becoming higher-order conscious of the fact that one has a particular lower-order state. Hence Lurz's interpretation of the HOT theory must be incorrect and his argument against the HOT theory must be rejected. Let's turn to my argument now.

6.1 The Higher-Order Thought Theory *Cannot* Characterize Phenomenal Character in terms of Higher-Order Consciousness

My argument is actually rather simple: Because the HOTs required for basic state consciousness are themselves unconscious states, we cannot characterize the phenomenal character of basic state consciousness in terms of one's coming to be conscious of what these HOTs represent. Let's lay out the facts that lead to this conclusion.

First, we agreed earlier that in attributing state consciousness to a particular mental state, we also must be identifying that there is some kind of phenomenal character affiliated with the instantiation of that state. Conversely, in identifying a mental state as *un*conscious, we must be identifying that there is not any phenomenal character affiliated with being in that mental state.

Second, HOT theorists are very clear that the second-order thoughts required for basic state consciousness are, themselves, unconscious mental states. We saw this, for example, in the way HOT theorists account for the seeming immediacy of the awareness we have of our mental states and in the fact that, according to the HOT theory, when a HOT itself *is* state conscious, what results is actually *introspective* state consciousness,

rather than *basic* state consciousness. Clearly, then, since the second-order thoughts required for basic state consciousness are themselves unconscious states, it therefore also must be the case that there cannot be any phenomenal character affiliated with the instantiation of these HOTs. With these facts in mind, let's take a fresh look at Lurz's interpretation of the HOT theory's account of phenomenal character.

Lurz takes the HOT theory to be claiming that the phenomenal character affiliated with *basic* state consciousness necessarily involves one's becoming higher-order conscious of the fact that one is instantiating the first-order state conscious state. Notice, however, that the fact that one is said to become conscious of here (the fact that one is in a particular first-order state) is precisely the fact that an appropriate HOT is presumed to represent. Recall, for example, that in order for George's belief that there is an apple on the table to be state conscious in the basic sense, George must form an appropriate HOT with the content, roughly, "I believe that there is an apple on the table". So George's HOT represents the fact that he instantiates this particular belief, and this is precisely the fact Lurz interprets the HOT theory to be claiming that George must become conscious of when his belief is state conscious in the basic sense. This means that, on Lurz's interpretation, the HOT theory claims that one becomes conscious of exactly what one's HOT represents, whenever one has a state conscious state, in the basic sense.

Notice, however, that in saying that it's the content of one's HOT that suddenly and immediately comes before one's mind, it would seem that one is saying that there *is* phenomenal character affiliated with instantiating these HOTs after all. I see no way of getting around this conclusion. In saying that there *is* phenomenal character affiliated with the HOTs required for basic state consciousness, however, one would thereby

contradict the HOT theory's very clear commitment to the fact that the HOTs involved in basic state consciousness are themselves unconscious states. Since unconscious states can have no affiliated phenomenal character, it cannot be the case both that these states are unconscious and that these states have affiliated phenomenal character; one of these claims must be rejected.

As it turns out, it appears that the HOT theory cannot reject the claim that the HOTs involved in basic state consciousness are unconscious states, however. If HOT theorists were to reject this claim, and hence were to say that these HOTs are themselves state conscious, then the theory would require that these HOTs be represented by yet higher-order states, and then those further higher-order states would also have to be conscious and so would also have to be represented by yet higher-order states, and so on. This reasoning would lead the theory into infinite regress. It would appear that the only option then is to reject the alternative claim, namely, that the HOTs involved in basic state consciousness have affiliated phenomenal character.

For these reasons, I take it the HOT theory *cannot* be claiming that the phenomenal character affiliated with basic state consciousness necessarily involves one's becoming higher-order conscious of the fact that one has a particular first-order mental state. If HOT theorists were to claim that basic state consciousness involves coming to be conscious of what one's *HOTs* represent, then they would be claiming that there *is* something it's like for one to be instantiating these HOTs, hence they would be contradicting their own claim that these HOTs are not state conscious. Since it appears that they cannot reject the claim that these HOTs are unconscious, we must instead reject the claim that there is a phenomenal character affiliated with instantiating these HOTs.

So we can conclude that the HOT theory *cannot* be claiming that one becomes conscious of the content represented by one's unconscious HOTs when one has mental states that are state conscious in the basic sense. It turns out, therefore, that Lurz must have mischaracterized the HOT theory's account of phenomenal character.

If the argument so far is sound, we can also conclude that Lurz's argument against the HOT theory must be rejected. Insofar as Lurz's argument is based on the assumption that HOT theorists account for the phenomenal character affiliated with basic state consciousness in terms of *higher-order* consciousness of what one's HOTs represent, and insofar as we now see that this cannot be the case (because the requisite HOTs are unconscious states), we can reject the assumption represented as premise 5 in my reconstruction of Lurz's argument, and thus we can reject Lurz's conclusion that the HOT theory's account of phenomenal character is mistaken.

6.2 Two Accounts Contained Within the Higher-Order Thought Theory

The discussion above brings to light an interesting fact about the HOT theory that often appears to go unnoticed. In light of the conclusions drawn above we can see that the HOT theory can and must provide distinct answers to the two questions identified earlier – the question of what constitutes state consciousness and the question of what it's like to have state conscious states. Specifically, with regard to the question of what grounds or constitutes state consciousness, HOT theorists clearly provide an answer in terms of higher-order representation, as they say that state consciousness *consists* in one's higher-order representing oneself as being in a certain lower-order mental state. As regards the question of phenomenal character, however, HOT theorists *cannot* be providing an answer in terms of higher-order representation, because they *cannot* be

saying that the phenomenal character affiliated with basic state consciousness involves one's becoming conscious of what one's unconscious HOT represents. Since only one of these questions is answered in terms of higher-order representation, we see that the questions of constitution and of phenomenal character must be separated if we are to fully comprehend the HOT account of state consciousness. We can also see now that it is Lurz's failure to notice the separation of these two accounts within the HOT theory that constitutes the mistake in his argument.

7. An Alternative Interpretation: The Higher-Order Thought Theory on the *Phenomenal Character* of State Consciousness

Though I will not provide a detailed discussion of this point here, I would like to note that I think it is actually possible for the HOT theorists to endorse an account of the phenomenal character affiliated with basic state consciousness that is similar to the same-order account that Lurz (2003, 2006) endorses. In fact, I even think there is textual evidence to support an interpretation of the theory as providing such an account. I will briefly outline how I think this new account might proceed, then I will present the textual evidence that I think supports this interpretation.

I think HOT theorists could argue that what it's like for someone who has a state conscious state in the basic sense is that she comes to be conscious of what her state conscious state represents. In order to see how HOT theorists might be endorsing such a view, however, we must introduce a distinction between full-blown *consciousness* and mere *awareness*. In particular, if we reserve the term 'consciousness' for all and only those phenomena affiliated with some sort of what-it's-like-ness, and introduce the term 'awareness' to identify any sort of mental registration of information that has no affiliated

what-it's-like-ness (for example, the kind of mental registration that occurs in subliminal perception or blindsight) then I believe the HOT theory can comfortably endorse the following two principles:

> **The Awareness Principle:** Let *M* be any mental state that at times can be state conscious and at other times can be unconscious. When a subject instantiates *M* without also forming an appropriate HOT about being in *M*, *M* will enable the subject to be aware of what *M* represents, in the sense that she can be differentially responsive to what *M* represents, but since *M* will not be state conscious, there will not be anything it's like for the subject to be instantiating *M* at that time.
>
> **The Phenomenal Character Principle**: When a mental state, *M*, is represented by an appropriate HOT, and hence when *M* is state conscious, the creature instantiating these states will become same-order conscious of what *M* represents.

Combined with the HOT theory's account of what constitutes state consciousness, these principles would entail the following characterization of the phenomenal character affiliated with state consciousness: A person who has a state conscious mental state in the *basic* sense, will become *conscious* of what her first-order mental state represents and, since she forms a second-order thought of the right sort but that second-order thought is unconscious, she will become *aware* of what her second-order thought represents. So, for example, if George instantiates the belief that there is an apple on the table, and George also forms an appropriate HOT that he has such a belief, he will come to be *conscious* of the fact that there is an apple on the table (or, as Lurz argues, of the proposition 'that there is an apple on the table') and he will come to be *aware* of himself as having that belief.

With these two principles, we can also draw out a full characterization of the phenomenal character affiliated with *introspective* state consciousness. Since introspective state consciousness involves forming a third-order thought that, in turn,

makes it the case that one's second-order thought is state conscious, we can see that a person with an introspectively state conscious mental state will become *conscious* of what her second-order thought represents and she will become *aware* of what her third-order thought represents. So, in terms of our example, George will become *conscious* of himself as believing that there is an apple on the table (and thereby also become *conscious* of the fact that there is an apple on the table or of the proposition 'that there is an apple on the table') and he will become *aware* of himself as thinking that he believes that there is an apple on the table (since this is what he represents with his unconscious third-order thought).[12]

And thus we see a new way of interpreting the HOT theory's account of the phenomenal character affiliated with basic and introspective state consciousness. Though, admittedly, HOT theorists are rarely careful to draw the distinction between awareness and consciousness that I've introduced here, they do sometimes draw this distinction and, when they do, their account does seem to be in line with the account I have just outlined. For example, consider Rosenthal's comments in the following passage (and please forgive the length of the quote, I was compelled to include it all because it captures a lot of the points I have argued for here):

> There is a natural way of understanding how conscious states differ from mental states that are not conscious. No mental state is conscious if the individual that is in that state is in no way aware of it. If somebody thinks, desires or feels something but is wholly unaware of doing so, then that thought, desire or feeling is not a conscious state.
>
> Experimental work on non-conscious perception typically exploits this commonsense observation. Participants sometimes deny seeing a stimulus even when there is evidence, say from priming, that the relevant

[12] Notice, on this account, one does becomes higher-order consciousness of what one's HOTs represent but only when one has states that are *introspectively* state conscious. This is in contrast to Lurz's interpretation of the HOT theory, where this sort of phenomenal character was presumed to be a necessary component of the phenomenal character affiliated with *basic* state consciousness.

> visual information has affected psychological processing. The effect on subsequent psychological processing, moreover, typically reflects the perceptual discriminations that are characteristic of conscious visual states, e.g. among colours and shapes. We commonly conclude that the visual state occurred but without being conscious. In such cases, a participant's denial of seeing the stimulus reflects not a failure to see, but simply a lack of awareness of seeing. Things are the same outside experimental work. If a person denies wanting something but acts as people typically do when they want that thing, then we see the person as having that desire, though a desire that is not conscious. Novelists and dramatists have described such situations for centuries.
>
> Higher-order theories take this commonsense observation as basic to understanding how conscious states differ from mental states that are not conscious. Because no mental state of which one is wholly unaware is conscious, conscious states are mental states we are in some suitable way aware of. Higher-order theories differ among themselves about just what kind of awareness is required for a mental state to be conscious, but they are agreed that a state's being conscious involves some form of HOA [i.e., Higher-Order Awareness].
>
> When somebody perceives something subliminally, so that the perception is not conscious, there is nonetheless a kind of awareness of the perceived stimulus. It may sound awkward to speak of a non-conscious state that nonetheless makes one aware of something, but we can distinguish the conscious and non-conscious cases in a completely natural way. When one subliminally perceives something, one is aware of that thing but not consciously aware of it; when one consciously perceives the stimulus, one is consciously aware of it.
>
> (Rosenthal (2012), p. 1425.)

Here we see Rosenthal's commitment to the same sort of distinction between consciousness and awareness that I introduced above. We also see him saying, in this admittedly rare instance where he draws such a distinction, that state consciousness consists in one's awareness of one's mental states, not in one's consciousness of those states. Finally, and most importantly, we see Rosenthal saying that what one is conscious *of* when one has a state conscious mental state is actually what the *lower-order* state represents, he does not say, and so Lurz was mistaken to interpret the HOT theorists as saying, that one becomes conscious of what one's *HOT itself* represents. I take this to be demonstrated in the last sentence quoted here, where Rosenthal says that subliminal

perception leads to a mere awareness of the stimuli, whereas conscious perception leads to a full-blown consciousness *of the stimuli.* Notice, Rosenthal does *not* say that conscious perception leads to a full-blown consciousness *of the fact that one has a particular perception*. In light of these comments it would appear that the interpretation of the theory I've just presented seems to fit well with Rosenthal's own more careful presentation of the HOT theory.

As I mentioned earlier, there is currently very limited discussion of the HOT theory's account of phenomenal character. Furthermore, because HOT theorists rarely do phrase things in a way that respects the distinction between consciousness and awareness that I've introduced above, they actually do sometimes speak in ways that seem more in line with Lurz's interpretation of the theory. It is for these reasons that I call other philosophers to join me in this discussion of the HOT theory's account of phenomenal character. I maintain that HOT theorists have always been very clear in noting that the HOTs required for basic state consciousness are, themselves, not state conscious states. So I maintain that it is this aspect of their theory, in conjunction with the assumption that there cannot be anything it's like to be in a mental state that is not state conscious, that forces us to reinterpret the explicit claims often made by HOT theorists and to conclude that in fact the theory *cannot* provide the account of phenomenal character that Lurz attributes to it. Though I think my arguments here are sound, I believe this is an interesting and underexplored area of the HOT theory that deserves more attention.

8. Conclusion

To recap, I have defended the HOT Theory of Consciousness against Lurz's Phenomenal Character Argument by arguing that the HOT theory *cannot* claim that the

phenomenal character affiliated with *basic* state consciousness involves one's becoming conscious of what one's unconscious HOTs represent. Since this conclusion entails that Lurz's argument rests on a mischaracterization of the HOT theory's account of phenomenal character, I can reject Lurz's argument.

This discussion also brought to light an important and often unnoticed aspect of the HOT theory, namely, that HOT theory provides distinct accounts of what grounds state consciousness and of what it's like to have state conscious states. Importantly, we learned that only the former account can be given in terms of higher-order representation. This is a lesson that we must take seriously if we are to fully comprehend and properly assess the HOT theory's account of state consciousness.

References

Block, N. (1995) "On a Confusion About the Function of Consciousness", *Behavioral and Brain Sciences, 18(2)*, p. 227-247.

Block, N. (2011) "The Higher Order Approach to Consciousness Is Defunct", *Analysis, 71(3)*, p. 419-431.

Dretske, F. (1995) *Naturalizing the Mind*, Cambridge, MA: MIT Press.

Lurz, R. W. (2003) "Advancing the Debate Between HOT and FO Accounts of Consciousness" *Journal of Philosophical Research, 28*, p. 23-44.

Lurz, R. W. (2006) "Conscious Beliefs and Desires: A Same-Order Approach" In Uriah Kriegel and Kenneth Williford (Eds.) *Self-Representational Approaches to Consciousness*, Cambridge, MA: MIT Press, p. 321-351.

Nagel, T. (1974) "What Is It Like to Be a Bat?", *Philosophical Review, 83(4*), p. 435-450.

Rosenthal, D. M. (2000) "Consciousness, Content, and Metacognitive Judgments", *Consciousness and Cognition*, *9*, p. 203-214.

Rosenthal, D. M. (2002a) "A Theory of Consciousness", In Ned Block, Owen Flanagan, and Güven Güzeldere (Eds.) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press, p. 729-753.

Rosenthal, D. M. (2002b) "Explaining Consciousness", In David J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford: Oxford University Press, p. 406-421.

Rosenthal, D. M. (2003) “Unity of Consciousness and the Self”, *Proceedings of the Aristotelian Society*, *103*, p. 325-352.

Rosenthal, D. M. (2004) “Varieties of Higher-Order Theory”, In Rocco J. Gennaro (Ed.) *Higher-Order Theories of Consciousness: An Anthology*, Amsterdam: John Benjamins Publishing Company, p. 17-44.

Rosenthal, D. M. (2005) *Consciousness and Mind*, Oxford: Clarendon Press.

Rosenthal, D. M. (2011) “Exaggerated Reports: Reply to Block” *Analysis, 71(3)*, p. 431-437.

Rosenthal, D. M., (2012) “Higher-Order Awareness, Misrepresentation and Function”, *Philosophical Transactions of the Royal Society B*, *367*, p. 1424-1438.

Weisberg, J. (2011) “Abusing the Notion of What-It’s-Like-Ness: A Response to Block” *Analysis, 71(3)*, p. 438-443

Paper 3: ***An Alternative Way to Defend the Higher-Order Thought Theory of Consciousness Against the Misrepresentation Objection***

1. Introduction

In this paper I present a new way of defending the Higher-Order Thought Theory of Consciousness against the Misrepresentation Objection.[1]

As it is usually understood, the HOT theory claims that a mental state is conscious if and only if it is represented by an appropriate higher-order thought (HOT) to the effect that one is in that very state. The theory allows, however, that a HOT might *mis*represent a person as being in a state that she is not in fact in and that, despite this misrepresentation, it would still seem to the person that she is in the state she, in fact, is not instantiating. Such cases of misrepresentation are known in the literature as Empty HOT cases and they are the source of a particularly persistent objection to the HOT theory – the Misrepresentation Objection.

According to Ned Block (2011a, 2011b), who is the most recent proponent of this objection[2], the problem is that the HOT theory explains conscious states by providing a set of conditions that are supposed to be necessary and sufficient for a person's having conscious states. Block argues that these conditions are shown to be inadequate, however, because the conditions lead to incompatible predictions about the presence of

---

[1] Though my claim is novel relative to anything David Rosenthal has ever said in response to the Misrepresentation Objection, it has recently come to my attention that Richard Brown (2012) may suggest a solution to this objection that is somewhat similar to my own. Specifically, like myself, Brown suggests that there may be two separate sorts of consciousness explained by the HOT theory. In contrast to my claim, however, it seems that both sorts of consciousness Brown discusses may be forms of state consciousness, insofar as they both are taken to be properties of mental states. I will not get into a detailed comparison of our accounts here, though I refer the interested reader to Brown (2012).

[2] Others have raised versions of the Misrepresentation Objection as well. For example, see Byrne (1997), Neander (1998), and Levine (2001).

conscious states in empty HOT cases. In light of this result, Block concludes that we cannot accept the explanation of conscious states that the HOT theory provides and, since the HOT theory's explanation fails, so the HOT theory must be rejected.

David Rosenthal is the main proponent of the HOT theory and he does provide his own reply to this Misrepresentation Objection.[3] His strategy is to argue that the objectors have misunderstood what it means to say that a mental state is conscious and hence that the objection, which is based on this false understanding of the theory, can be rejected. For example, Rosenthal often speaks as though a mental state's consciousness is a relational property, a property that consists in the relationship of a mental state's being represented by an appropriate HOT. Despite often accounting for a mental state's consciousness in this way, however, when responding to the Misrepresentation Objection Rosenthal consistently makes comments such as the following:

> Since there can be something it's like for one to be in a state with particular mental qualities even if no such state occurs, a mental state's being conscious is not strictly speaking a relational property of that state. A state's being conscious consists in its being a state one is conscious of oneself as being in. Still, it's convenient to speak loosely of the property of a state's being conscious as relational so as to stress that it is in any case not an intrinsic property of mental states.
>
> (Rosenthal (2005), p. 211.)

Here, Rosenthal suggests that a mental state's consciousness is *not* actually a relational property after all, instead conscious mental states are simply the mental states *we are aware of ourselves as being in*. Since one can be *aware of oneself as* being in a particular mental state even if one is in no such state, the issue of whether or not a lower-order state is actually instantiated becomes irrelevant to the theory's explanation of this sort of awareness, and hence irrelevant to the theory's account of mental state

[3] See, for example, Rosenthal (2004, 2011).

consciousness. This entails that empty HOT cases, cases where no relevant lower-order state is instantiated, pose no threat to the HOT theory after all.

Though I do believe that Rosenthal's reply succeeds in addressing the Misrepresentation Objection, I also believe that the objectors find it unconvincing. This is evidenced, for example, by the fact that objectors continue to reissue their attack, despite Rosenthal's consistent responses each time the attack is issued. It seems that objectors have trouble accepting Rosenthal's reply precisely because his strategy involves clarifying the meaning of key terms and concepts within the theory. Though Rosenthal does introduce these clarifications in some of his earliest writings about the HOT theory[4], most discussions of the theory are still couched in terms of the loose, relational notion of mental state consciousness. This means that the details of the more strict account, and the implications that account has for other aspects of the theory, are rarely if ever fully explored. Rosenthal's strategy of responding by appealing to this clarified account therefore leaves the objectors feeling slighted; they feel suspicious of whether the clarifications Rosenthal suggests are really as minor as he would have us believe. In light of these facts, it is hardly surprising that objectors need further convincing before they are willing to accept that the objection has been defeated.

In this paper I take on the task of convincing the objectors by presenting an independent reply to the Misrepresentation Objection. What's unique about my strategy is that I try to maintain the terms of the debate that the objectors are familiar with, in hopes of presenting a reply that the objectors will find more convincing. Though I do think my solution arrives at a final characterization of the HOT theory that is similar to

[4] See, for example, Rosenthal (1986).

the one Rosenthal suggests with his own solution, I will not engage in any argument to that effect here. (Nor will I discuss Rosenthal's reply any further in this paper.) Instead, my goal is just to provide a response to the Misrepresentation Objection that more clearly identifies where the objection departs from the HOT theory and hence that allows the objectors to understand how the reply to the objection is satisfying.

To provide a quick preview, first I argue that Block's Misrepresentation Objection can be seen as being based on an equivocation on two senses of the phrase 'conscious state'. Then I argue that the HOT theory can be seen as providing separate explanations for each sense of 'conscious state' and that neither of these explanations turn out to have internally incompatible necessary and sufficient conditions in the empty HOT cases. Since neither explanation turns out to have internally incompatible conditions, I conclude that Block's Misrepresentation Objection can be rejected.

The paper proceeds as follows: I begin by briefly introducing the most common interpretation of Rosenthal's HOT theory and outlining Block's (2011a, 2011b) version of the Misrepresentation Objection. Because my goal is to present a reply in terms that the objectors agree upon, it would be unhelpful for me to describe the HOT theory in the more careful terms Rosenthal uses in his reply to the Misrepresentation Objection. Despite that fact, note that when I say that a particular claim is made by *Rosenthal* himself, that claim is one that I believe Rosenthal would endorse, even in his most careful or strict presentations of his theory. On the other hand, when I say that a claim is made by *HOT theorists*, that claim is likely only part of the loose characterization of the theory, i.e., the characterization of the theory that is most often presented.

After setting out this common understanding of the debate, I then introduce two senses of the phrase 'conscious state', I demonstrate how Block's representation of the HOT theory's necessary and sufficient conditions can be seen as confusing these two senses, and I draw three lessons: First, I argue that the HOT theory can be seen as offering a separate explanation of each sort of consciousness identified by the different senses of 'conscious state'. Second, I argue that neither of these explanations has internally incompatible necessary and sufficient conditions in empty HOT cases, hence I conclude that we can reject Block's objection. Finally, I briefly discuss one implication of my new interpretation of the HOT theory – the fact that it entails that there can be something it's like for a person with empty HOTs, despite the fact that no mental state is state conscious.

With our plan now in place, let's begin by outlining the HOT theory and the Misrepresentation Objection.

2. The Common Understanding of the Higher-Order Thought Theory of Consciousness

I will here introduce the HOT theory as it is most often presented and, importantly, as the objectors understand it. This will enable us to draw out the Misrepresentation Objection and, eventually, to resolve the objection in a way that objectors should find acceptable.

The HOT theory is most commonly presented as a theory of state consciousness. State consciousness is the property that a mental state is said to have when it's conscious and to lack when it's unconscious. For example, George may consciously desire a drink of water and hence may set out to quench his thirst. On his way to the faucet, however,

perhaps the phone rings and George gets caught up in a conversation with his wife. Though George still has the desire for a drink of water at this point, perhaps that desire now ceases to be conscious. In this example, George's desire would change from being *state conscious* at one moment to being *unconscious* at the next.

In saying that a mental state is state conscious, HOT theorists are taken to be saying that there's something it's like for the person instantiating that state to be having that state at that time. For example, in saying George's desire is state conscious, HOT theorists would be saying that there's something it's like for George to be consciously desiring a drink of water, and that there ceases to be anything it's like for George to desire the water when that desire ceases to be state conscious. HOT theorists are therefore taken to be providing not just an account of state consciousness but also an account of there coming to be something it's like for a person with a state conscious mental state.

The account of state consciousness provided by HOT theorists is initially grounded on an insight borrowed from the everyday folk: We find it natural to say that if a person is not aware of her mental state in any way, then her mental state is not conscious. This implies, conversely, that a conscious mental state must be a state a person is aware of being in. HOT theorists take this folk intuition to entail that the difference in a mental state when it's conscious and when it's unconscious lies in its possessor's awareness of it. And so HOT theorists set out to explain state consciousness by explaining how a person comes to be aware of her mental states. As you might have guessed, HOT theorists explain how we come to be aware of our mental states in terms of our forming higher-order thoughts about those mental states. Specifically, the HOT

theory is commonly understood to be arguing that a mental state is state conscious if and only if that mental state is represented by an appropriate[5] higher-order thought.

Note that, insofar as state consciousness is understood to consist in a person's awareness of her mental states, and insofar as this awareness is realized by a HOT representing that lower-order state, state consciousness is understood to be a *relational* property of mental states; a property grounded in the representational relationship between a HOT and the lower-order mental state it represents. Since this is the characterization of state consciousness most familiar to objectors, this is the characterization of state consciousness that we will work with throughout the paper.

So far then, there are five important facts about the theory that we must be sure to keep in mind throughout the paper: First, that the HOT theory is taken to be providing an explanation of state consciousness. Second, that state consciousness is taken to be a property attributed to mental states when there's something it's like for the bearer of those states to be instantiating those states. Third, that we are understanding state consciousness to be a relational property of mental states. Fourth, that the HOT theory's explanation of state consciousness is provided in terms of a person's coming to be aware of her mental states. And finally, that a person is said to become aware of her mental states by forming an appropriate higher-order thought about herself as being in those states.

---

[5] The use of the word 'appropriate' here can be important since there are some conditions a HOT must meet if it is to be *appropriate*. For example, the HOT must have an assertoric mental attitude and it must not be the product of any conscious inferences or observations. As these conditions will not play a role in my argument, however, I will not discuss them further here. For a concise discussion of the conditions that make a HOT appropriate as well as the arguments leading to the HOT theory of state consciousness presented here see Rosenthal (2002), especially Section II, "The Hypothesis", pg. 408-411.

To bring all of this together then, the HOT theory is saying, for example, that in order for George's desire for a drink of water to be state conscious, George would need to form an appropriate HOT with the content, roughly, "I desire a drink of water". His desire and this HOT would thus stand in the appropriate relation, there would come to be something it's like for George to have that desire, and George's desire would thereby count as being state conscious. With this understanding of the basic aspects of the theory now in place, we can move to a discussion of the aspect of the theory that is specifically targeted by the Misrepresentation Objection.

Unlike some other variations of higher-order theories[6], Rosenthal is pretty clear that he thinks it is possible for a higher-order thought to *mis*represent its bearer's mental environment. To demonstrate this possibility of misrepresentation, and to draw out the key fact about the theory that opens the door to the Misrepresentation Objection, let's consider three sample cases.

In the first case, a person might, for example, see a green apple. This would lead her to form a first-order visual state with the content 'green apple'. If the subject forms an appropriate HOT about this visual state, perhaps with the content, roughly, 'I see a green apple', then the subject's visual state will be state conscious. Since the subject's HOT accurately represents her mental environment, this is a case of *Veridical HOT Representation*.

In the second case, let's suppose instead that our subject sees a red apple and so she forms a first-order visual state with the content 'red apple'. According to Rosenthal,

---

[6] Though Rosenthal argues that a HOT is a mental state that is distinct from the lower-order state it represents, both Gennaro's (1996) Wide Intrinsicality View and Van Gulick's (2004) Higher-Order Global States View posit that the HOT is actually a component, along with the lower-order state it represents, in a larger, more complex mental state. It might be argued, therefore, that such views can avoid the misrepresentation problems facing Rosenthal's HOT account.

it's completely possible that, despite having this first-order visual state of seeing a red apple, our subject nonetheless might form an appropriate[7] HOT with the content 'I see a green apple'.[8] In this case, the subject's HOT mischaracterizes her lower-order visual state as being about a green instead of a red apple, so this is a case of *Misrepresentation*. As we will see, this form of misrepresentation is comparatively mild however, since there is a lower-order state instantiated by the individual, the problem is just that it is mischaracterized by her HOT. For this reason, these sorts of cases are referred to as cases of *Mild HOT Misrepresentation*.

Finally, in the third case, we might suppose that our subject does not see any apples whatsoever, and hence that she forms no relevant first-order visual states at all. Rosenthal allows that even in this case our subject still might form an appropriate HOT with the content 'I see a green apple'. Since there is no relevant visual state in these cases at all, however, the HOT is not just mischaracterizing the nature of the state the subject is in but rather is mischaracterizing the subject as being in any such relevant lower-order state in the first place. This is why these cases are known as cases of *Radical HOT Misrepresentation* or as *Empty HOT* cases; they are referred to as *empty* HOT cases precisely because the lower-order state the HOT represents the subject as being in is not in fact instantiated by the subject.

Now, Rosenthal not only allows that all three cases are possibilities within the theory, but he also says that what it's like for a person will be the same in all three cases.

---

[7] Notice that by calling these HOTs appropriate, I am *not* saying that they are accurate or veridical representations. Instead, I intend the word 'appropriate' to identify that the HOTs in question meet the various conditions Rosenthal sets out for the HOTs involved in state consciousness. (See footnote 5 for further discussion).

[8] For a sample of Rosenthal's views on the possibility of misrepresenting HOT cases see Rosenthal (2004), especially pg. 32-35 and Rosenthal (2005), especially pg. 217-218.

Specifically, since in all three cases the subject formed a HOT representing herself as seeing a green apple, Rosenthal says that what it's like for the subject will be as though she actually is seeing a green apple in each of these cases. Rosenthal comes to this conclusion because he argues that the way a HOT characterizes one's mental environment will determine what it's like for a person to be in a conscious state. He says this explicitly, for example, when he writes, "…what it's like for one will follow the way one's HOT represents one's state, even when that HOT misrepresents the state one is actually in" (Rosenthal (2005, p. 217).[9] Because Rosenthal is committed to there being something it's like for the subject in all three cases however, he also appears to be committed to saying that the subject is indeed in a *conscious state* in each of these cases. It is precisely this fact that leads Block to raise the Misrepresentation Objection.

3. Block's Misrepresentation Objection

Block (2011a, 2011b) objects to the HOT theory by arguing that the conditions the theory says are necessary and sufficient for conscious states actually lead to conflicting predictions about whether someone is in a conscious state or not in empty HOT cases. In light of this fact, Block argues that the theory's explanation of conscious states fails and so the HOT theory should be rejected.

To demonstrate this, Block summarizes the theory's necessary and sufficient conditions as follows: "…an appropriate higher order thought is *sufficient* for a conscious

[9] Rosenthal also writes, "If I consciously take something I see to be a cow when it's actually a horse, phenomenologically it's as though I consciously see a cow. Similarly, if I am conscious of myself as being in a *P* state, it's phenomenologically as though I'm in such a state whether or not I am. If I'm not in a *P state*, that will make a difference to my overall mental functioning… But the phenomenology is determined solely by the way I am aware of things…" (Rosenthal (2004), p. 35). Here we see not only Rosenthal's claim that the phenomenology resulting from veridical and misrepresenting HOTs will be subjectively indistinguishable, but also his suggestion that these subjectively indistinguishable differences might be detectable by other means (for example, they might make a difference to one's overall mental functioning). For further discussion of this point see Rosenthal (2005), p. 29-30 and Rosenthal (2011).

state and…being the object of an appropriate higher order thought is *necessary* for a conscious state" (Block (2011b), p. 443, original emphasis). Block then argues that in empty HOT cases, such as the third case described above, it's clear that the condition sufficient for there to be a conscious state is satisfied. Since there is an appropriate HOT formed in the empty HOT scenario, and since the presence of such a HOT is sufficient for there to be a conscious state, the theory dictates that there will be a conscious state in the empty HOT scenario. As we discussed, Rosenthal says there will be something it's like for the subject in this case - it will seem to her that she is seeing a green apple - so Rosenthal seems to agree that the subject will indeed be in a conscious state in such cases. The trouble, argues Block, is that the condition necessary for there to be a conscious state is, at the same time, not satisfied.

Recall that, according to Block, the necessary condition dictates that a mental state must be the object of an appropriate HOT if there is to be a conscious state. In the empty HOT scenario, however, this is precisely what is stipulated *not* to be the case. Since the HOT is empty, there is no actually instantiated first-order state that can be said to be the state that the HOT represents. Barring a dramatic shift in the theory to something more like a self-representational theory (i.e., to a theory wherein the HOT somehow represents itself[10]), there is thus no candidate for being the state that is represented and hence no candidate for being the state that is conscious. Thus the condition necessary for there to be a conscious state fails to be satisfied in empty HOT cases and hence the theory also predicts that there will *not* be a conscious state in these cases.

[10] For accounts from philosophers endorsing this sort of view, and their discussions of how accepting this view allows them to avoid the Misrepresentation Objection, see Gennaro (2012), especially Ch. 4 and Kriegel (2003), especially pg. 119-120. Obviously Rosenthal does not endorse a self-representational view.

As it turns out then, in empty HOT cases the condition sufficient for there to be a conscious state is met while, at the same time, the condition necessary for there to be a conscious state is not satisfied. The theory therefore predicts both that there will and that there will not be a conscious state in these cases. Therein lies Block's problem with the theory. Since the HOT theory explains conscious states by providing conditions that are supposed to be necessary and sufficient for conscious states, and since those conditions turn out to be incompatible in empty HOT cases, Block argues that the HOT theory cannot have provided an adequate explanation of conscious states after all. He concludes that the theory therefore fails and should be rejected.

To summarize the concern slightly differently, the problem boils down to the following: When a HOT is empty, and thus when there is no instantiated lower-order mental state that can be said to be the apparently necessary object of the HOT, it seems that there also must be no candidate lower-order state that can be said to be the state that is conscious. However, Rosenthal clearly allows that there is something it's like for the bearer of empty HOTs, thus he appears to allow that there is indeed a conscious state in such cases. The problem then becomes one of reconciling these two facts – the fact that there is something it's like for the bearer of an empty HOT and the fact that there is no lower-order state that can be said to be the conscious state the bearer of that HOT is instantiating. Thus we are left with an apparent tension within the theory. We are now ready to see my solution to this problem.

4. Replying to the Misrepresentation Objection

I propose that we can reconcile the apparent tension in the HOT theory by arguing that each of the conflicting conditions is actually a condition for a separate sort of

consciousness. To my knowledge, Rosenthal has never endorsed such a view but, as we'll briefly discuss at the end of the paper, I think that support for this claim can be found within the HOT theory. My argument in support of this solution begins by identifying the fact that the phrase 'conscious state' is ambiguous, so let's turn to that step now.

4.1 The Ambiguity of the Phrase 'Conscious State'

I submit that the phrase 'conscious state' is ambiguous. People commonly use one sense of the phrase 'conscious state' wherein this phrase describes a mental state *that is* conscious. However, and importantly, people also commonly use another sense of the phrase 'conscious state' wherein the phrase describes a creature's state *of being* conscious. For example, someone might say that George *is in a conscious state*. In saying this, the person might mean that a particular mental state of George's, his desire for a drink of water, say, is conscious and hence that George has a mental state *that is* conscious. On the other hand, the person might mean that, insofar as George is consciously desiring things like drinks of water and is not passed out due to dehydration, George himself is in a state *of being* conscious.[11] Let's examine the differences between these two claims a little further.

The first sense of 'conscious state' identified here, according to which the phrase describes a mental state *that is* conscious, matches up with what HOT theorists are traditionally taken to be identifying with their notion of *state consciousness*. As the HOT

[11] Those immersed in debates about the HOT theory might recognize the distinction drawn here as potentially mirroring the common distinction between intransitive state consciousness on the one hand and transitive and intransitive creature consciousness on the other. Though I am not convinced the notion of subject consciousness that I introduce here will map neatly onto the usual notions of creature consciousness, I choose to avoid the standard terminology primarily in hopes of minimizing theoretical baggage and maximizing clarity.

theorists are most often understood to use this term (and I will share in this usage), state consciousness is a property attributed to *mental states* and it is the property that differentiates mental states that there's something it's like for their bearer to be instantiating from mental states that there's nothing it's like for their bearers to be instantiating. For example, before George was distracted by the phone call from his wife, his desire for a drink of water was a *state conscious* desire and there was something it was like for George to have that desire.

The second sense of 'conscious state' identified here is the sense captured by saying that a creature is in a state *of being* conscious. I propose that this sense of 'conscious state' identifies a property attributed to *subjects* (i.e., to people or other suitable creatures), rather than a property attributed to mental states. I also propose that we can understand this property as being the property that differentiates creatures for whom there's something it's like from those for whom there's not something it's like.[12] So, for example, George might be said to have this property both before and during his phone call with his wife. Before the call, what it's like for George might be characterized by saying that he wanted a drink of water. During the phone call, however, when George's desire for water was no longer immediately before his mind, there would be some other way of characterizing what it's like for George to be chatting with his wife. If there was nothing it was like for George, however – if he was a philosophical zombie, for

[12] By using the phrase 'something it's like' here, I intend to loosely identify the aspect of consciousness that I believe Thomas Nagel (1974) famously discusses. This feature, according to Nagel, is something subjective and experiential, something it's like *for an organism* to be conscious or to have a conscious mental state. This essential, subjective nature of consciousness is also sometimes referred to as the *qualitative* or *phenomenal* character of consciousness or simply as *what-it's-like-ness*. The reader should note that I intend no theoretical ties to any one particular account of this sort of consciousness when I use these terms though. Instead, I just mean to identify a more intuitive notion, similar to what Weisberg (2011) calls the "moderate" reading of these terms. Furthermore, since I attribute this property to *subjects* rather than *mental states*, it's not clear my notion of this property can be mapped onto the usual accounts anyway.

example – then George would lack the sort of property identified by this sense of 'conscious state' entirely. I propose that we call this property *subject consciousness*, since this sort of property is only attributed to subjects, not to mental states. (After all, while we are fine saying that there's something it's like *for a subject*, we would find it quite odd to say that there's something it's like *for a mental state*.) So here we've characterized the *subject consciousness* sense of 'conscious state'.

To recap, we have identified two separate senses of the phrase 'conscious state' – a state consciousness sense and a subject consciousness sense. When we use 'conscious state' in the *state consciousness* sense we mean to identify a property attributed to *mental states* that consists in the relation between those mental states and the appropriate HOTs that represent them. We also intend to identify mental states for which there's something it's like for their bearers to be in them. On the other hand, when we use 'conscious state' in the *subject consciousness* sense of the phrase we mean to identify a property attributed to a *subject* and a property that differentiates subjects for whom there is something it's like from subjects for whom there is nothing it's like.

Having identified these two different senses of the phrase 'conscious state', we are now ready to see how Block's representation of the HOT theory's necessary and sufficient conditions can be seen to equivocate on these two senses of 'conscious state'.

4.2 Block's Inadvertent Equivocation on 'Conscious State'

Recall, Block represented the HOT theory as endorsing the following necessary and sufficient conditions for conscious states: First, that "being the object of an appropriate higher order thought is *necessary* for a conscious state" and second, that "an appropriate higher order thought is *sufficient* for a conscious state" (Block (2011b), p.

443, original emphasis). Notice that each of Block's conditions is supposed to be either necessary or sufficient for a *conscious state*. To determine which sense of 'conscious state' is at play in each condition, let's look at each in turn.

As for the necessary condition, I think it's clear that this condition must be taken to be using the *state consciousness* sense of 'conscious state' for the following reasons: First, as we saw, Rosenthal explicitly claims that there *will* be something it's like for a creature in empty HOT cases, and these cases count as *empty* HOT cases precisely because there is no mental state that is the object of the instantiated HOT. In light of these facts, the HOT theory cannot be taken as saying that it is a necessary condition for subject consciousness (i.e., for there being something it's like for a subject) that there be a mental state that is the object of a HOT. Such an interpretation would contradict Rosenthal's claims and so would not be a fair way of interpreting the HOT theory. Hence we can conclude that this necessary condition cannot be interpreted as using the subject consciousness sense of 'conscious state'.

Secondly, this necessary condition is taken to be a part of the common way of understanding the HOT theory's explanation of state consciousness. As we saw earlier, HOT theorists are understood to be claiming that a mental state is state conscious if and only if it is represented by an appropriate HOT. Block's necessary condition (that being the object of an appropriate HOT is necessary for a conscious state) clearly fits into this account of state consciousness. So, since the condition does not cohere with Rosenthal's comments about subject consciousness, and since it does fit with the HOT theory's common account of state consciousness, it appears we must conclude that this necessary condition uses the phrase 'conscious state' in the *state consciousness* sense of that phrase.

As for the sufficient condition, "that an appropriate HOT is sufficient for a conscious state", I think the most natural way to understand this condition is to see it as using 'conscious state' in the *subject consciousness* sense of that phrase. My reasoning here is as follows: First and, again, as we learned from Rosenthal's comments on the empty HOT scenarios, the theory *does* dictate that whenever there is an appropriate HOT, there will be something it's like for the creature who has that HOT. Thus the HOT theory does appear to endorse the claim that having an appropriate HOT is sufficient for subject consciousness, in my sense of that term.

Secondly, I believe we cannot make sense of this condition as applying to state consciousness without eliminating the very possibility of there being empty HOTs and hence without contradicting Rosenthal's own claims that these scenarios are possible within the HOT theory. For example, if we were to rephrase Block's sufficient condition such that it explicitly refers to state consciousness, then we might end up with the following condition: "An appropriate HOT is sufficient for a lower-order mental state that is state conscious". Notice how this condition appears to suggest that a lower-order state is guaranteed to be instantiated whenever a person also instantiates an appropriate HOT. We could make this result even more explicit by representing the condition as follows: "If a lower-order mental state is instantiated and there is an appropriate HOT formed about it, then that lower-order mental state is state conscious".

If we rephrase the condition in either of these ways, however, we appear to have eliminated the very possibility that a HOT can be empty in the first place. After all, the fact that the lower-order state is instantiated is now built right into these versions of the sufficient condition, if not explicitly then implicitly. So, in trying to interpret this

condition as being about state consciousness, it turns out that we are actually forced to deny that the HOT theory allows for the possibility of empty HOT cases. As we saw, however, Rosenthal does explicitly allow that empty HOT cases are possible within the theory. Since it turns out, then, that the sufficient conditions resulting from a state consciousness interpretation of 'conscious state' are in direct conflict with Rosenthal's claims about the theory, we must conclude that interpreting this sufficient condition to be using the state consciousness sense of 'conscious state' is not a fair way of representing the HOT theory.

Instead, given that Rosenthal intends the theory to allow for the possibility of empty HOT scenarios and given that Rosenthal does say that one's having a HOT is sufficient for there being something it's like for one in such cases, I conclude that the most appropriate interpretation of the sufficient condition is therefore the one that takes it to be using the *subject consciousness* sense of 'conscious state'.

To bring these insights together now, we've seen that each of Block's conditions actually seems to be a condition for a different sort of consciousness. His necessary condition seems to be identifying something that's necessary for *state* consciousness, whereas his sufficient condition seems to be identifying something that's sufficient for *subject* consciousness. Once the phrase 'conscious state' is disambiguated in these ways, we can reconstruct Block's statement of the necessary and sufficient conditions as follows: A creature's having an appropriate HOT is sufficient for a creature's being *subject conscious* and a mental state's being the object of an appropriate HOT is necessary for the mental state to be *state conscious*. With these clarifications in place we

can see clearly now that Block's presentation of the theory's necessary and sufficient conditions does involve a problematic equivocation on the phrase 'conscious state'.

4.3 Three Lessons

There are three lessons we can draw from the above analysis. Let's look at each in turn.

4.3.1. The First Lesson – Two Explanations Within the Higher-Order Thought Theory

The first lesson is that the HOT theory can now be seen as providing two separate explanations, one for each sense of 'conscious state' that was identified.[13] Let's see how each explanation can be drawn out of the facts that we have established already.

First, since we have established that a mental state's being the object of an appropriate HOT is taken to be a necessary condition for state consciousness, we can hold that condition constant and draw out the following explanation of state consciousness:

> **State Consciousness**
> If a mental state, *x*, is the object of an appropriate HOT, then *x* is state conscious.
> If a mental state, *x*, is state conscious, then *x* is the object of an appropriate HOT.

Hence, a mental state's being the object of an appropriate HOT is both necessary and sufficient for that mental state to be state conscious.

Second, since we established that a creature's having an appropriate HOT can be taken to be a sufficient condition for that creature's being subject conscious, we can hold that condition constant and draw out the following explanation of subject consciousness:

---

[13] Again, note that Rosenthal has never claimed that his account provides two separate explanations. As I argue later in the paper, however, I believe there is support for my account within his theory.

Subject Consciousness
If a creature, *y*, is having an appropriate HOT, then *y* is being subject conscious.
If a creature, *y*, is being subject conscious, then *y* is having an appropriate HOT.

Hence, a creature's having an appropriate HOT is both necessary and sufficient for the creature's being subject conscious.

Thus we can see that there are indeed two explanations that can be provided by the HOT theory – one explanation of state consciousness and one explanation of subject consciousness.

4.3.2 The Second Lesson – No Internal Incompatibility

The second lesson is that once we've identified these different explanations, we can see that neither has conditions that are internally incompatible in the empty HOT scenario. (This result is summarized in Table 3 and I will refer to that table throughout the analysis.)

We can see in the first row of Table 3 Block's original statement of the HOT theory's necessary and sufficient conditions. In each of the second and third rows I have provided a reconstruction of the actual necessary and sufficient conditions that the HOT theory can be seen as providing for state consciousness and subject consciousness respectively. Notice that Block's ambiguous phrase 'conscious state' (represented in bold in row 1) is disambiguated in each new condition (as shown in the italicized parts of rows 2 and 3). Also notice that the condition we found to be appropriate for each sort of consciousness is held constant within the appropriate row (as highlighted by asterisks).

Table 3. *Clarifying the necessary and sufficient conditions provided by the Higher-Order Thought Theory*.

	Necessary Condition	Sufficient Condition	Result in Empty HOT Scenario
Block's Representation of the HOT Theory's Conditions	"being the object of an appropriate higher order thought is necessary for a **conscious state**"	"[having] an appropriate higher order thought is sufficient for a **conscious state**"	There is no object of the HOT so the necessary condition is not satisfied. The subject does have an appropriate HOT, so the sufficient condition is satisfied. The two conditions lead to differing predictions about the presence of a conscious state in empty HOT cases - there is an incompatibility among these conditions.
State Consciousness Sense of 'Conscious State'	A mental state's being the object of an appropriate HOT is necessary for *the mental state to be state conscious*	*A mental state's being the object of an appropriate HOT* is sufficient for *the mental state to be state conscious*	There is no state that is the object of the HOT, so neither condition will be satisfied. Both conditions predict that there will not be a state conscious state in empty HOT cases - there is no incompatibility among these conditions.
Subject Consciousness Sense of 'Conscious State'	*A creature's having an appropriate HOT* is necessary for *the creature's being subject conscious*	A creature's having an appropriate HOT is sufficient for *the creature's being subject conscious*	There is a creature with an appropriate HOT, so both conditions will be satisfied. Both conditions predict that there creature will be subject conscious in empty HOT cases - there is no incompatibility among these conditions.

With all these clarifications in place, we can now see that the new conditions for each sort of consciousness are not internally incompatible in the empty HOT case after all. Let's look at the results for state consciousness first.

The newly clarified conditions for state consciousness are as follows: A mental state's being the object of an appropriate HOT is necessary for *the mental state to be state conscious* and *a mental state's being the object of an appropriate HOT* is sufficient for *the mental state to be state conscious*.

We can now see that, rather than there being an incompatibility of predicted outcomes here, the conditions for state consciousness (as we and the objectors understand that property) just fail to be satisfied in the empty HOT scenario. That is, since there is no state that is the object of an appropriate HOT and since this condition is both necessary and sufficient for state consciousness, the theory, as I've interpreted it, predicts that there will not be any state conscious mental state in empty HOT cases. Block's charge of an incompatibility among necessary and sufficient conditions therefore fails to be realized here.

Notice that, since the very circumstances of an empty HOT scenario dictate that there is no lower-order state present, and hence that there is no lower-order state to which we could attribute the property of state consciousness in the first place, this result is not problematic for the theory. When a HOT is empty, it turns out to be true both that there is no state that is the object of the HOT and also that there is no state to which we can attribute the property of state consciousness anyway. Hence the result that there is no state conscious state in these cases should not be alarming.

As for subject consciousness, the newly clarified conditions are as follows: *A creature's having an appropriate HOT* is necessary for *the creature's being subject conscious* and a creature's having an appropriate HOT is sufficient for *the creature's being subject conscious*.

Since these conditions make no mention of the lower-order state that fails to be instantiated in empty HOT cases, we can see that these conditions are both satisfied and, as Rosenthal has maintained, the theory therefore predicts that there will be something it's like for the creature in the empty HOT cases. Furthermore, since the necessary and sufficient conditions for subject consciousness are here both satisfied, there turns out not to be any internal incompatibility among the conditions for subject consciousness in empty HOT cases either. Block's charge of an internal incompatibility among conditions again fails to be realized.

In light of these findings, we can conclude that in empty HOT cases there is no internal incompatibility among the conditions we've now taken the HOT theory to provide, either for state consciousness or for subject consciousness. The conditions for state consciousness both fail to be met in empty HOT scenarios and the conditions for subject consciousness are both satisfied in empty HOT scenarios. From here we can conclude that Block's charge that the HOT theory has internally incompatible conditions in empty HOT cases is actually based on his accidental combination of one condition from each of the two separate explanations. Since the HOT theory provides no explanation consisting of those two conditions together, and hence the theory has no explanation with internally incompatible conditions, we can safely reject Block's Misrepresentation Objection.

This clarification of the explanations provided by the HOT theory not only eliminates the objection but also helps us to understand how Block might have fallen into endorsing this objection in the first place. The most obvious reason for this mistake is clearly the fact that Rosenthal has never said that the theory provides two separate explanations, hence Block can hardly be faulted for failing to notice that the two conditions he identifies are part of two different explanations contained within the theory.

Even more interestingly, however, we now also can see how Block's mistake of running together a condition from each of the different explanations would be an easy mistake to make, *even if* he was aware of the presence of the two separate explanations. The reason is that both of these separate explanations actually take the formation of an appropriate HOT to be a necessary condition for their particular sort of consciousness. Specifically, mental states are state conscious when appropriate HOTs are formed about them, so a HOT must be formed in order for any mental state to be state conscious, and creatures are subject conscious when they are representing via HOTs, so a HOT also must be formed in order for any creature to be subject conscious. Given this parallel in the theory's explanations of the two sorts of consciousness, we can see clearly now that it would be rather easy to confuse the conditions involved in the two explanations.

4.3.3 The Third Lesson – Subject Consciousness Without State Consciousness

And now, with the objection safely eliminated, we can finally come to our third lesson. We can see now that what is, in fact, the case in empty HOT scenarios is that no mental state is state conscious, yet there's still something it's like for the creature with the empty HOT; the creature with the empty HOT is still subject conscious, in my sense of that term. Since we've separated the notions of state and subject consciousness, this

result is not problematic for the theory. The theory can simply be understood as dictating that the subject of an empty HOT can be subject conscious without, at the same time, having any individual mental states that are state conscious.

In coming to this conclusion, we also have finally identified the point at which the objection departs from the HOT theory. Objectors appear to have been taking state consciousness to play a bigger role in determining the presence and character of subject consciousness than the HOT theory ends up allowing.[14] Though on our understanding of the HOT theory, part of what we mean when we say a mental state is state conscious is that there is something it's like for the bearer of that state to be instantiating that state at that time, it turns out that what it's like for the subject, the character of her subject consciousness, is not directly determined by her state conscious state after all. Again, it now seems obvious that this would have to be the case, since a subject can have subject consciousness without having any relevant state conscious mental state at all.

Furthermore, we've seen that there is support for this interpretation in Rosenthal's own comments, as he says that there will be something it's like for a creature with empty HOTs and that what it's like for this creature will be determined by the way its HOTs represent the creature's mental environment. In saying these things, Rosenthal also appears to be claiming that lower-order states (those states represented by HOTs) have no *direct* role to play in grounding what it's like for a creature in empty HOT scenarios. I take it, therefore, that he is drawing a very similar conclusion to the one that I am drawing when I say that creatures can be subject conscious without having any mental states that are state conscious. So Rosenthal and I appear to end up with converging

[14] This mistake is abundantly clear, for example, in the form of the Misrepresentation Objection raised by Neander (1998) and Levine (2001).

accounts of the HOT theory after all, despite arriving at our conclusions by means of different response strategies.[15]

For HOT theorists, then, there are a few interesting outcomes of the reply I've presented here, which I'd like to take a moment to note. First, though it might be surprising to say that a creature can have subject consciousness without having any mental states that are state conscious, this claim appears to be in line with Rosenthal's more careful comments about the HOT theory. Perhaps we can take it as an added benefit of my solution, then, that it brings to light an important and underexplored aspect of the HOT theory of consciousness. I predict that there is fruitful research to be done exploring this consequence further. HOT theorists owe us a more detailed account of the connections (and disconnections) between state and subject consciousness.

Second, the fact that the theory has this surprising result about the independence of subject consciousness from state consciousness is not itself any reason to reject the HOT theory. The objectors may not be convinced yet that this account is correct, but they no longer have any specific reason to reject the theory on this basis. So we have succeeded in defeating the Misrepresentation Objection. Furthermore, since we've accomplished all this with a discussion presented in the terms objectors are familiar with, I hope that we've reached these conclusions in a way that objectors also will find more convincing.

---

[15] Though Rosenthal and I allow that there might be no *direct* role for state conscious states to play in determining phenomenal character, it still might be the case that these states do play a necessary though *indirect* role in accounts of HOT representation and hence in accounts of subject consciousness. For example, a HOT might not be able to *mis*represent someone as being in a state of pain if it weren't for the fact that the person was also able to form an accurate representation of herself as being in pain. Though I note this complication for the reader, I do not intend to follow up on this point here.

5. Conclusion

To recap, we defeated the Misrepresentation Objection by first disambiguating the phrase 'conscious state' in order to identify two sorts of consciousness, each with its own separate explanation provided by the HOT theory. We then saw that, in empty HOT cases, neither of these separate explanations have internally incompatible conditions. Instead, it turned out that the Misrepresentation Objection was based on Block's inadvertently running together a condition from each of the two separate explanations. We were thereby able to reject the Misrepresentation Objection. Finally, we saw that one consequence of interpreting the HOT theory in the manner suggested here was that the HOT theory must now be taken as denying any direct role for first-order states in grounding the phenomenal character of conscious experience (at least in empty HOT cases). We also saw, however, that, though controversial, this consequence both is predicted by Rosenthal and is not in itself a legitimate reason to reject the HOT theory. Thus the HOT theory was successfully defended against the Misrepresentation Objection and this defense was conducted in terms that the objector should find convincing.

References

Block, N. (2011a) "The Higher Order Approach to Consciousness Is Defunct", *Analysis, 71(3)*, p. 419-431.

Block, N. (2011b) "Response to Rosenthal and Weisberg" *Analysis, 71(3)*, p. 443-448.

Brown, R. (2012) "The Brain and Its States", In Shimon Edelman, Tomer Fekete, Neta Zach (Eds.), *Being in Time: Dynamical Models of Phenomenal Experience*, Amsterdam: John Benjamins Publishing Company, p. 211-230.

Byrne, A. (1997) "Some Like It HOT: Consciousness and Higher Order Thoughts", *Philosophical Studies, 86*, p. 103-129.

Gennaro, R. J. (1996) *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*, Amsterdam: John Benjamins Publishing Company.

Gennaro, R. J. (2012) *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts.* Cambridge, MA: MIT Press.

Kriegel, U. (2003) "Consciousness as Intransitive Self-Consciousness: Two Views and an Argument", *Canadian Journal of Philosophy, 33(1)*, p. 103-132.

Levine, J. (2001) *Purple Haze: The Puzzle of Consciousness*, Oxford: Oxford University Press.

Nagel, T. (1974) "What Is It Like to Be a Bat?", *Philosophical Review, 83(4*), p. 435-450.

Neander, K. (1998) "The Division of Phenomenal Labor: A Problem for Representational Theories of Consciousness", *Philosophical Perspectives, (12), p. 411-434.*

Rosenthal, D. M. (1986) "Two Concepts of Consciousness", *Philosophical Studies, 491,* p. 329-359.

Rosenthal, D. M. (2002) "Explaining Consciousness", In David J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford: Oxford University Press, p. 406-421.

Rosenthal, D. M. (2004) "Varieties of Higher-Order Theory", In Rocco J. Gennaro (Ed.) *Higher-Order Theories of Consciousness: An Anthology*, Amsterdam: John Benjamins Publishing Company, p. 17-44.

Rosenthal, D. M. (2005) *Consciousness and Mind*, Oxford: Clarendon Press.

Rosenthal, D. M. (2011) "Exaggerated Reports: Reply to Block" *Analysis, 71(3)*, p. 431-437.

Van Gulick, R. (2004) "Higher-Order Global States (HOGS): An Alternative Higher-Order Model of Consciousness", In Rocco J. Gennaro (Ed.) *Higher-Order Theories of Consciousness: An Anthology*, Amsterdam: John Benjamins Publishing Company, p. 67-92.

Weisberg, J. (2011) "Abusing the Notion of What-It's-Like-Ness: A Response to Block" *Analysis, 71(3)*, p. 438-443.

Conclusion

To conclude this dissertation, I'd like to present some reflections on the discussions just provided. The main lessons of this dissertation are, I believe, identified by the very title of the project: *The Resilience of a Refined Higher-Order Thought Theory of Consciousness*.

The reference to *resilience* in the title reflects the fact that this dissertation consists of three independent papers, each defending the HOT Theory of Consciousness against a different objection. As we saw, I defeat the Theory of Mind objection by demonstrating that the metarepresentational capacities measured by Theory of Mind tests are not parallel to the metarepresentational capacities required by the HOT theory for the formation of HOTs, hence that the objection, which is based on assuming these requirements are parallel, can be rejected.

I defeat the Phenomenal Character Argument by showing that the HOT theory cannot provide an account of phenomenal character (or subject consciousness) that requires a person to become suddenly and immediately *conscious* of what her *HOTs* represent. Hence this objection, which is based on the assumption that HOT theorists do provide such an account, can be rejected.

Finally, I defeat the Misrepresentation Objection by demonstrating that the HOT theory actually contains two different explanations of two different sorts of consciousness. Once these explanations are separated, we see that the objection is actually based on a confusion about the relationship between these two accounts, rather

than being based on any incompatibility within either account, hence I show that this objection also can be rejected.

Insofar as I am able to defeat these objections, then, I take it that I have shown the HOT theory to be resilient.

Another key word in the title is '*refined'*. Most discussions of the HOT theory are focused on its account of mental state consciousness – the property that differentiates mental states when they are conscious, from mental states when they are not conscious. Though I do not deny that this is an important aspect of the theory, my arguments throughout the dissertation suggest that there is *another* important aspect to the HOT theory, one that seems to go largely unnoticed in critical discussions. As we've seen, this second aspect is an account of *what it's like for a person* – an account of phenomenal character or subject consciousness. I argue that this second account is importantly distinct from the theory's account of mental state consciousness and that this account of subject consciousness appears to have significant explanatory power.

To see how this side of the project is realized consider, for example, that we learn in the second paper that the character of phenomenal experience, literally *what* it is like for a person, can be captured in terms of one's coming to have, immediately before one's mind, the content of a represented lower-order state[1], rather than the content of a HOT itself. This shows us that the character of subject consciousness is not explained in terms higher-order representation and hence, that the account of subject consciousness might be importantly distinct from any account the theory provides of state consciousness.

[1] Strictly speaking, we might say that what comes immediately before one's mind is actually the content that a HOT represents a lower-order state as having, as phrasing things in this more precise way allows us to capture the character of the subject consciousness afforded in Empty HOT scenarios.

We see a similar lesson in the third paper as well, where we learn that the refined HOT theory allows for the possibility that a person might be subject conscious without also having a mental state that is state conscious. This is further evidence that state consciousness and subject consciousness are importantly distinct sorts of consciousness and also that the explanations offered of these two sorts of consciousness are importantly distinct within the theory.

Finally, we see the potential explanatory power of such an account of subject consciousness at work in the first paper, since it is only by bringing to light the Phenomenal Character Principle and the Awareness Principle, both of which are aspects of the refined HOT theory's account of subject consciousness, that we are able to demonstrate why the evidence from TOM tests is not able to support the TOM objection.

Hence, the theory that is shown to be resilient, insofar as it is capable of defeating these objections, is also found to be a new and refined version of the HOT theory of consciousness.

No philosophical discussion seems to be complete if it fails to offer up new questions for future research. Thankfully, it seems there are many new questions raised by this project and I'd like to conclude this discussion by highlighting one such question in particular.[2] Though I cannot provide a complete response to this question here, as doing so would be the foundation of yet another research paper, I will make a few comments in response to this concern, in hopes of helping us reflect a little more on some

[2] Thank you to Professor William Seager for bringing this concern to my attention (in a commentary on an earlier draft of the third paper, presented at the Canadian Philosophical Association Conference in May 2012) and to Professor Andrew Brook for also raising this concern (in comments on an earlier draft of this dissertation).

of the ideas introduced in this work and in hopes of suggesting potential avenues for further inquiry. So here is the question:

> *Is the account of consciousness put forth in this dissertation still a* higher-order *account of consciousness?*

Let's first look a little further into the motivation for such a concern.

The HOT theory originally might have been understood as an attempt to explain consciousness in terms of the structural relations between some mental states and other thoughts that represent those mental states. Furthermore, such an understanding might have suggested that we identify certain states as being *higher-order* states in the first place, precisely because they instantiate the right sorts of structural relationships with other, lower-order mental states. Given this way of understanding the theory, however, there ought to be no thoughts that count as *higher-order* thoughts, and there ought to be no explanation of consciousness available either, when it turns out that no lower-order state is instantiated.

Since we learn, in the third paper of this dissertation, that the refined HOT theory allows one to have subject consciousness without this structural relation holding (i.e., without there being a relation between one's higher-order thought and a lower-order state, because no lower-order state is instantiated), then it might seem both that some sort of consciousness is being explained without appeal to the structural relationship that appeared to be necessary for a higher-order explanation of consciousness and, furthermore, that certain thoughts are being identified as *higher-order* thoughts while failing to stand in the appropriate structural relationships to other mental states.

In short, if we agree that an explanation is provided in *higher-order* terms only if it appeals to the right structural relations among mental states, and if we agree that a thought counts as a *higher-order* thought only if it stands in the right structural relation to other mental states, then the refined higher-order thought theory presented in this dissertation appears to have given up the game of providing a *higher-order* account of consciousness.

Here is a brief sketch of one possible reply to this concern: Perhaps state consciousness traditionally has been explained by appeal to this sort of structural relation. Though Rosenthal (2005) suggests that he never actually intended for state consciousness to be a relational property of mental states (instead he says that any reference to this property as being relational was just a useful shorthand[3]), I have been happy, in this dissertation, to leave that common interpretation of the theory intact and hence to maintain that state consciousness really is this sort of relational property. That is why, in the third paper, I come to the conclusion that there is no instantiation of state consciousness in empty HOT scenarios. Since there is no relation instantiated, because there is no lower-order state to comprise one of the relata in this relation (and also no lower-order state to instantiate the property of state consciousness), I argue that there will fail to be any mental state that is state conscious in empty HOT scenarios. Hence my account of state consciousness should not be in jeopardy of loosing its status as a *higher-*

[3] For example, we see this when Rosenthal writes, "Since there can be something it's like for one to be in a state with particular mental qualities even if no such state occurs, a mental state's being conscious is not strictly speaking a relational property of that state. A state's being conscious consists in its being a state one is conscious of oneself as being in. Still, it's convenient to speak loosely of the property of a state's being conscious as relational so as to stress that it is in any case not an intrinsic property of mental states" (Rosenthal (2005), p. 211).

order account, because it does fall in line with the traditional way of identifying higher-order accounts in terms of their appeal to structural relations among mental states.

On the other hand, I have also suggested that another sort of consciousness, subject consciousness, *is* present in these empty HOT scenarios, and hence is present when the apparently necessary relations are not instantiated. In fact, I have argued explicitly that subject consciousness is not a relational property of mental states, but rather it is a property of subjects – people or other suitable creatures – themselves. The account of subject consciousness that emerges from my dissertation is one in which a subject's higher-order representing, of herself as being in certain mental states, is both necessary and sufficient for her being subject conscious. So it is not a relation between mental states that matters here, but rather a particular sort of representing.

But now we return to the second side of this question – if the representations involved do not stand in the appropriate structural relations to other mental states, how can they count as being *higher-order* representations? And if they do not count as being higher-order representations, how can my explanation count as a *higher-order* explanation of subject consciousness?

My response to this worry is as follows: We can identify certain thoughts as being *higher-order* thoughts by appealing to their content, rather than appealing to their extrinsic structural relations. That is, thoughts will be identified as being higher-order thoughts only if they are representations of other mental states (and hence only if they involve the activation of concepts for other mental states like BELIEVE, DOUBT, DESIRE, SMELL, and so on).[4]

[4] Note, my claim here need not be that a thought counts as a higher-order thought only if it involves the very concepts of belief, doubt, and so on that we use in our western adult folk-psychology. There is room

Furthermore, to be the kinds of higher-order thoughts that figure in the explanations offered here, these higher-order thoughts must also meet Rosenthal's (2002) further conditions on appropriateness. That is, they must be noninferential, nondispositional, assertoric thoughts to the effect that one is, oneself, in a particular mental state.[5]

Finally, since one can token a thought about one's mental environment even though that thought is not a veridical representation of that environment, in the same way that one can token a thought about one's external environment even though that thought is not a veridical representation of the external environment, we can see that it is possible to identify thoughts as being *higher-order* thoughts, on this characterization, without there being a lower-order thought instantiated that bears the right structural relation to the higher-order thought.

In short, if we type thoughts by means of their content (or by means of their activation of certain concepts) then we can identify certain mental states as *higher-order* states, without concern for the structural relations they actually stand in. This means, therefore, that the account of subject consciousness offered here would count as a *higher-order* explanation of subject consciousness, precisely because the thoughts that figure as necessary components of this explanation do in fact count as being *higher-order* representations. Again, the account of subject consciousness that emerges from my dissertation is one in which a subject's higher-order representing, of herself as being in certain mental states, is both necessary and sufficient for her being subject conscious.

---

here for allowing that other, perhaps less rich, concepts for mental state types might be sufficient, and presumably certain creatures or even young human infants might make use of these other concepts.

[5] For a concise discussion of the conditions that make a HOT appropriate, see Rosenthal (2002), especially Section II: "The Hypothesis", pg. 408-411.

And the subject's act of representing counts as an act of *higher-order* representing, precisely because it involves the activation of concepts for other mental states.

Now, as we saw in the first paper of the dissertation, this metarepresentational requirement, the requirement that a subject be able to represent her other mental states, may be a feat that some subjects are unable to accomplish. This would mean that those subjects could not form the HOTs necessary for both state consciousness and subject consciousness, hence this would mean that there should not be anything it's like for such subjects. As I also argued in the first paper, the evidence we have currently, from comparative and developmental psychology, is inconclusive as to whether any creatures fail to meet these metarepresentational requirements, so we currently do not have any reason to suspect that this aspect of the theory is problematic.

Note, however, that the question of whether or not a creature can meet these metarepresentational requirements *is* an empirical question, and the fact that the refined HOT theory raises questions that can be tested empirically is actually a fact that bodes well for the theory. If we find empirical evidence that suggests creatures can be subject conscious without forming the right higher-order thoughts, the theory would be falsified. I consider it to be an indication of the strength of the current theory that it makes empirical predictions and that it is susceptible to empirical falsification.

To sum up, then, though we might care less about the name we use for the account and more about the account's potential to get us a better understanding of consciousness and to provide us with empirically testable hypotheses, one might resist the urge to abandon all ties to the Higher-Order family of accounts just yet for the following reasons: It still seems rather important, even on the refined account presented

here, that both sorts of consciousness I have distinguished still require HOT formation. That is, no existing lower-order mental state will instantiate the relational property of state consciousness unless its bearer forms an appropriate HOT about herself as being in that state. Furthermore, no person will be subject conscious unless that person forms an appropriate HOT about herself as being in some other mental state (regardless of whether that higher-order representation is veridical). Since we can identify thoughts as being *higher-order* thoughts on the basis of their content (or on the basis of the concepts they require a creature to activate), rather than having to rely on the actual structural relations these thoughts bear to other mental states, we can see that the thoughts involved in this account can truly count as being *higher-order* thoughts, and we can maintain that the account of consciousness offered here is truly a *higher-order* account of consciousness.

As mentioned in some of the papers, there are still many other questions this account raises as well. For example, since we've distinguished two sorts of consciousness we can now ask what the relationship is between subject consciousness and state consciousness and what kinds of roles each sort of consciousness plays for an organism. We might also ask what sorts of entities our refined HOT theory would posit as the reduction base of each of these separate sorts of properties. Finally, and importantly, we might explore the implications these accounts have for the empirical work currently being conducted. Have the ideas presented here gotten us closer to answering questions about the neural correlates of consciousness? Or have they suggested other ways in which it might be better to understand the relationship between our minds and our brains? Has the analysis offered in the first paper shed any light on

experimental procedures that might help us to better assess the metarepresentational capacities of subjects in theory of mind tests?

Though I will not attempt to answer these questions here, it is my hope that the account I have offered sparks curiosity in the minds of fellow philosophers and psychologists, and that questions such as the ones identified here will now take centre stage in future research on the (refined) Higher-Order Thought Theory of Consciousness.

References

Rosenthal, D. M. (2002) "Explaining Consciousness", In David J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford: Oxford University Press, p. 406-421.

Rosenthal, D. M. (2005) *Consciousness and Mind*, Oxford: Clarendon Press.

Curriculum Vitae

Name

Lee-Anna T. Sangster

Education

The University of Western Ontario,
London, Ontario, Canada
2000-2004, B.A. Hon.,w/ Distinction, Philosophy & Psychology

Simon Fraser University
Burnaby, British Columbia, Canada
2004-2006, M.A., Philosophy

The University of Western Ontario,
London, Ontario, Canada
2006-2013, Ph.D., Philosophy

Awards

Western Scholarship of Distinction – 2000
University of Western Ontario In-Course Scholarship – 2002
Edna Jeffery Scholarship – 2002
Allison H. Johnson Scholarship in Philosophy – 2003
Simon Fraser University Special Graduate Entrance Scholarship – 2004
Social Sciences and Humanities Research Council (SSHRC) Canada Graduate Scholarship, Master's – 2004-2005
Simon Fraser University Graduate Fellowship (Master's) – 2005
Ontario Graduate Scholarship – 2007-2008; 2008-2009
Western Canadian Philosophical Association Student Essay Prize – 2008
Scholarship for First International Summer School in Cognitive Sciences and Semantics – 2009
Richard A. Harshman Scholarship – 2011
Western Graduate Thesis Research Award – 2011
Canadian Philosophical Association Student Essay Prize - First Place – 2012

Teaching Experience

Instructor: The University of Western Ontario, London, ON, Department of Philosophy, January-April 2011, September-December 2009.
Guest Lecturer: King's University College, London, ON, Department of

Psychology, March 2012; The University of Western Ontario, London, ON, Department of Philosophy, Feb 2012, November 2010.

Teaching Assistant: The University of Western Ontario, Department of Philosophy, January-April 2009, September 2007-April 2008, September 2006-April 2007; Simon Fraser University, Burnaby BC, Department of Philosophy, January-April 2006, January-April 2005, September-December 2004.

Grader: The University of Western Ontario, London, ON, Department of Philosophy, September-December 2008.