‘Consciousness,’ Consciousness, and Self-Consciousness

BY

JOSEPH N. GOTTLIEB

B.A., University of Southern California, 2007

THESIS

Submitted as partial fulfillment of the requirements for the degree of Doctor of Philosophy in Philosophy in the Graduate College of the University of Illinois at Chicago, 2016

Chicago, Illinois

Defense Committee:

David Hilbert, Chair & Advisor
Mahrad Almotahari
Marya Schechtman
David Rosenthal, City University of New York Graduate Center
Daniel Stoljar, Australian National University
ACKNOWLEDGEMENTS

It took a village to raise this dissertation.

First things first: I am lucky to have had a fantastic dissertation committee. My committee—comprised of Mahrad Almotahari, Marya Schechtman, David Rosenthal, and Daniel Stoljar, with Dave Hilbert quarterbacking—is about the best one could ask for. Mahrad has been a consistent source of advice on things related to the dissertation and otherwise, and in many ways saved the first portion of the dissertation from oblivion. Marya was initially one of the reasons I came to UIC to begin with, having wanted to work on issues at the intersection of consciousness and personal identity. Though I ended up directing my attention elsewhere, some of those ideas nonetheless snuck into the third part of the dissertation, and my conversations with her—beginning at the (sadly, now defunct) Bad Ass Coffee on the north side of Chicago over three years ago—helped tremendously. The influence of my two external committee members on my thinking is hard to overestimate. David’s work on consciousness is without a doubt the central animating force for much of what follows, and I thank him for lending me his ear and his timely feedback. Part of this dissertation would not have been possible without the theoretical tools supplied by Daniel’s own work, turning a hunch into a concrete argument. But Daniel has also been a source of encouragement and feedback beginning during my time at ANU and beyond, and I thank him for all his help.

The biggest thanks, though, goes to Dave, my chair. There is a reason I call Dave a ‘Swiss Army Knife’: I have never met anyone whose knowledge of philosophy has such breadth. But the analogy is far from perfect, because it’s not merely breadth: its depth, too. His ability to penetrate
deep into an issue, no matter the topic, and see where I had not, has been invaluable. He has endured many of my false starts, and showed me how to be patient and careful, and has been just an overall excellent model of a philosopher. For whatever philosophical talent I have gained over the years, I owe much of it to him.

Plenty of other people helped along the way. Colin Klein has been a source of support throughout my development as a philosopher, both during his time at UIC and elsewhere. I thank David Chalmers for sponsoring my trip to ANU during the (northern) summer of 2013, where a sizeable chunk of the ideas in this dissertation where born. ANU is an inspiring place to do philosophy. I also thank Richard Brown, Sam Coleman, Uriah Kriegel, and especially Jonathan Farrell and Jeff Speaks for helpful feedback. I thank Emily Gottlieb for doing much of the late-stage editing. Her relentless eye for detail was much appreciated. Josh Norton, Burkay Ozturk, and Saja Parvizian have been great friends throughout my time at UIC. I owe special thanks to Saja, both as a friend and as a collaborator. He provided comments on the entire dissertation, and our work together on Descartes has made me a better philosopher. I would be remiss not to also mention Nick Huggett as DGS, and Valerie Brown, our departmental secretary. Val: I am sorry for asking you for my copier code 10, 556 times in the past five-plus years.

Finally, I thank my wife Jessica and my son Theo. Words don’t do any justice, so I won’t offer any beyond this: you make the writing worth it. This dissertation is dedicated to them.

JNG
SUMMARY

My dissertation advances three primary theses, divided across three self-standing papers. The first two theses concern the Transitivity Principle, the claim that conscious mental states are mental states we are aware of in some way. The Transitivity Principle is the core of ‘Higher-Order’ approaches to consciousness. My first thesis is that the dispute between Higher-Order and ‘First-Order’ theorists (those who reject the Transitivity Principle) is verbal. But unlike paradigmatic verbal disputes, the dispute is substantive. So, to settle the dispute, we need to take a hard look at the language central to our discourse about consciousness.

Next, I look at a puzzle local to Higher-Order theories. Proponents of the Transitivity Principle do not just claim that the Transitivity Principle is true. They say something stronger: the Transitivity Principle is a truism. However, it is also plausibly a truism that experience is transparent is to awareness. Hence the puzzle: two independently plausible claims, both of which are central to the study of consciousness, are prima facie incompatible. It is uncontroversial that the best theory will, ceteris paribus, be the theory that has the resources to grant both claims. More controversial is my second thesis: there is no such theory.

My third and final thesis concerns the relationship between consciousness and self-consciousness. A central challenge to those who would argue that all consciousness involves self-consciousness is Hume’s well-known claim that the self is in some sense phenomenologically elusive. I articulate and defend a novel theory—neutral between Higher-Order and First-Order approaches—according to which all consciousness involves self-consciousness, but one that nonetheless accepts, rather than rejects, Hume’s claim.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td><strong>1  Four Theses in the Philosophy of Mind</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>2  Verbal Disputes in the Theory of Consciousness</strong></td>
<td>8</td>
</tr>
<tr>
<td>2.1 The Transitivity Principle</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Using and Abusing The Nagelian Conception of Consciousness</td>
<td>10</td>
</tr>
<tr>
<td>2.3 The Transitivity Dispute is a Verbal Dispute</td>
<td>14</td>
</tr>
<tr>
<td>2.3.1 Preliminaries</td>
<td>14</td>
</tr>
<tr>
<td>2.3.2 The Semantics of ‘What it is Like’-Sentences</td>
<td>20</td>
</tr>
<tr>
<td>2.3.3 Higher Order English and First Order English</td>
<td>23</td>
</tr>
<tr>
<td>2.3.4 Disputed and Undisputed Sentences, Take One</td>
<td>30</td>
</tr>
<tr>
<td>2.3.5 Disputed and Undisputed Sentences, Take Two</td>
<td>40</td>
</tr>
<tr>
<td>2.4 Prolegomenon to a Future Science of Consciousness</td>
<td>42</td>
</tr>
<tr>
<td><strong>3  Transitivity and Transparency</strong></td>
<td>48</td>
</tr>
<tr>
<td>3.1 Two Purported Truisms about Consciousness</td>
<td>48</td>
</tr>
<tr>
<td>3.2 Getting Transparent on Transparency</td>
<td>51</td>
</tr>
<tr>
<td>3.3 Acceptable Consistency Glosses</td>
<td>57</td>
</tr>
<tr>
<td>3.4 The Many Varieties of Transitivity</td>
<td>63</td>
</tr>
<tr>
<td>3.4.1 Unconscious Transitivity</td>
<td>63</td>
</tr>
<tr>
<td>3.4.2 Peripheral Transitivity</td>
<td>66</td>
</tr>
<tr>
<td>3.4.3 Non-State-Relational Transitivity</td>
<td>69</td>
</tr>
<tr>
<td>3.4.4 De Dicto Transitivity</td>
<td>71</td>
</tr>
<tr>
<td>3.4.5 Indirect Transitivity</td>
<td>75</td>
</tr>
<tr>
<td>3.5 Not So Truistic After All?</td>
<td>79</td>
</tr>
<tr>
<td><strong>4  Self-Consciousness Despite Elusiveness</strong></td>
<td>81</td>
</tr>
<tr>
<td>4.1 Subjectivist Theories of Consciousness</td>
<td>81</td>
</tr>
<tr>
<td>4.2 The Challenge from Elusiveness</td>
<td>84</td>
</tr>
<tr>
<td>4.2.1 Articulating Elusiveness</td>
<td>84</td>
</tr>
<tr>
<td>4.2.2 Motivating Elusiveness</td>
<td>89</td>
</tr>
<tr>
<td>4.2.3 The Object Constraint</td>
<td>94</td>
</tr>
<tr>
<td>4.3 Adverbialism about Self-Consciousness</td>
<td>98</td>
</tr>
<tr>
<td>4.4 The View</td>
<td>105</td>
</tr>
<tr>
<td>4.4.1 Lonergan on Self-Consciousness</td>
<td>105</td>
</tr>
<tr>
<td>4.4.2 Egocentric Representationalism</td>
<td>107</td>
</tr>
<tr>
<td>4.5 Elusiveness and Transparency</td>
<td>115</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Subjectivist Theories of Consciousness (4.1) 82
LIST OF FIGURES

1. Naturalness and Gerrymandered Predicates (2.3.1) 17
1 FOUR THESES IN THE PHILOSOPHY OF MIND

Thematically, this dissertation is organized around four theses central to the study of the nature of phenomenal consciousness, self-consciousness, and the relation between them.¹ The first thesis is TRANSITIVITY: conscious mental states are mental states we are aware of in some way. TRANSITIVITY is a thesis about the nature of conscious mental states:

Explaining what it is in virtue of which conscious states differ from mental states that aren’t conscious is the principle goal of a theory of consciousness. And it’s fairly straightforward to get a start on that question. When a mental state is conscious, the individual that’s in that conscious state is conscious of it; when a mental state fails to be conscious that individual is in no way conscious of it (Rosenthal 2005: 4).

TRANSITIVITY marks the central divide between the two main theoretical approaches to consciousness: first-order (FO) approaches and higher-order (HO) approaches. FO theories (e.g. Block 2007, 2008: 289; Dretske 1995; Tye 2000, 2009) reject TRANSITIVITY. HO theories (e.g. Lycan 2001; Rosenthal 2005; Kriegel 2009) accept TRANSITIVITY.

The second thesis is THE NAGELIAN CONCEPTION OF CONSCIOUSNESS, or NAGELIAN for short: if a subject is in a conscious mental state, then there is something it is like for that subject to be in that mental state. Like TRANSITIVITY, NAGELIAN is also a thesis about the nature of conscious mental states. NAGELIAN is so-called because Thomas Nagel was primarily responsible for its dissemination.² Here is how Nagel puts it:

[A]n organism has conscious mental states if and only if there is something that it is like to be that organism—something it is like for the organism (Nagel 1974: 436).

¹ Henceforth I will largely drop the ‘phenomenal’ qualification, and just use ‘consciousness’ as shorthand for ‘phenomenal consciousness.’ It will be clear from context when I depart from this usage.
² The namesake is only due to Nagel’s status as its popularizer; NAGELIAN, or something near enough like it, actually predates Nagel. See, e.g., Farrell (1950) and Sprigge (1971) and Wittgenstein (1974, § 91).
A passing glance at the philosophical literature on consciousness will show that, rightly or wrongly, this dissemination is *rife*; *NAGELIAN* is extremely popular.\(^3\) It is common ground for FO and HO theorists alike. For example, Ned Block (a FO theorist) tells us that:

Phenomenal consciousness is experience; what makes a state phenomenally conscious is that there is something “it is like” (Nagel, 1974) to be in that state (2002: 206).

And Michael Tye (another FO theorist) says:

A mental state, then, may be said to be phenomenally conscious just in case there is something it is like to undergo the state ... (1997: 290)

Likewise, according to David Rosenthal (a HO theorist):

When one lacks conscious access to a state, there is literally nothing it’s like for one to be in that state. Without access to a state one has no first-person perspective on it, and so there is nothing it’s like to be in it. As Thomas Nagel has insisted, what matters for consciousness is that there be something it’s like “for the organism” (Nagel, 1979, p.166). And there will be something it’s like for the organism only if the organism has conscious access to the relevant state (2000: 275).

And Uriah Kriegel (another HO theorist) says:

Phenomenal consciousness is the property mental states...have when, and only when, there is something it is like for their subject to undergo them, or be in them (2006: 58)

Thus while FO and HO theorists uniformly depart over *TRANSITIVITY*, FO and HO theorists are, to a significant extent, in agreement over *NAGELIAN*.

The third thesis is *TRANSPARENCY*. *TRANSPARENCY* concerns the phenomenology associated with conscious mental states. G.E. Moore was arguably the first to draw our attention to it:

[T]hat which makes the sensation of blue a mental fact seems to escape us; it seems, if I may use a metaphor, to be transparent—we look through it and see nothing but the blue (1903: 446).

---

\(^3\) Nagel’s thesis is slightly different from *NAGELIAN*. Nagel frames his thesis in terms of their being something it is like for a *creature* (or “organism”) to have conscious states. By contrast, *NAGELIAN* concerns what it is for a *mental state* to be conscious. Since the emphasis on conscious mental states is in line with the literature, I’ll ignore this quibble going forward. That said, state-consciousness and there being something it is like for a creature to be in a mental state are presumably connected (Kriegel 2009). See also Snowdon (2010: 11), Janzen (2011: 273), and Stoljar (forthcoming) for more on these points.
When we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous (1903: 450).  

Yet it was Gilbert Harman, not Moore, who really put the idea on the map:

When Eloise sees a tree before her, the colors she experiences are all experienced as features of the tree and its surroundings. None of them are experienced as intrinsic features of her experience. Nor does she experience any features of anything as intrinsic features of her experiences. And that is true of you too...Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict that you will find that the only features there to turn your attention to will be features of the tree... (1990: 667).

TRANSPARENCY is popular. Yet it is also tricky. For one thing, these quotes from Moore and Harman don’t obviously point to an unequivocal understanding of the same phenomenon. For another, it is prima facie simply odd to deny that we are aware of our experiences. Puzzlement having been registered, I propose we just leave things open-ended, with a promise to tighten up later: TRANSPARENCY says, roughly, that (i) there is some sense in which we are not aware of our experiences (or sensations, or conscious states) themselves, and that (ii) there is some sense in which we are only aware of the external ambient environment.

The fourth and final thesis is ELUSIVENESS. Like TRANSPARENCY, ELUSIVENESS also concerns the phenomenology associated with conscious states. Its most famous expression—at least in Western circles—comes from Hume’s Treatise:

For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other... I never can catch myself without a perception, and never can observe anything but the perception (1978: 252).

Like TRANSPARENCY, ELUSIVENESS is popular (e.g. Kant 1929; Ryle 1949; Shoemaker 1968; Howell 2010; Prinz 2012; Zahavi 2006a, 2014). And again like TRANSPARENCY, ELUSIVENESS is, well, elusive. Prima facie, if we are aware of anything, it is ourselves! Still, for now, I propose (again) that we

---

4 Although there are interpretive questions (irrelevant here) whether Moore’s considered position actually included a commitment TRANSPARENCY (Kind 2003).

FOUR THESSES IN THE PHILOSOPHY OF MIND
just leave things open-ended, with (again) a promise to tighten up later: ELUSIVENESS says, roughly, that there is some sense in which the self is not an item of conscious awareness.

Taken individually, each thesis—TRANSITIVITY, NAGELIAN, TRANSPARENCY, and ELUSIVENESS—has been discussed extensively in the literature. Yet relatively little has been said about the relation between them. I aim to remedy this situation.

Organizationally, the dissertation is divided into three self-standing papers. The first paper—“Verbal Disputes in the Theory of Consciousness”—concerns the dispute over TRANSITIVITY, and the relationship between TRANSITIVITY and NAGELIAN. I argue that the dispute over TRANSITIVITY is best understood as a verbal dispute. Specifically, I argue that the dispute over TRANSITIVITY is verbal because (i) FO theorists and HO theorists disagree about the meaning of the expression “for its subject” in NAGELIAN, and (ii) they would agree about TRANSITIVITY if they agreed on the meaning of the expression “for its subject” in NAGELIAN. The dispute over TRANSITIVITY, however, is not like other paradigmatic verbal disputes, which are also non-substantive. And because these disputes are non-substantive, when we realize this, we stop talking about the subject matter of the dispute. But this is not the right reaction in the case of the dispute over TRANSITIVITY. That’s because the dispute over TRANSITIVITY, though verbal, is substantive. The right reaction is to continue discussing TRANSITIVITY, but in a different way. To decide whether TRANSITIVITY is true, and by extension whether a FO or HO approach is correct, we need to look at the language central to our discourse about consciousness. And that means taking a hard look at NAGELIAN, and settling on a correct semantics for ‘what it is like’-sentences.

The second paper—“Transparency and Transitivity”—concerns the relationship between TRANSITIVITY and TRANSPARENCY. HO theories are largely individuated by how they understand
TRANSITIVITY’s ‘in some way’ clause—i.e. by what kind of higher-order awareness they claim is necessary for a mental state to be conscious. I’ll argue that there is no precisification of TRANSITIVITY’s ‘in some way’ clause that is (i) compatible with the various ways of understanding TRANSPARENCY yet (ii) continues to motivate HO theory.

This result is important. HO theorists don’t just say that TRANSITIVITY is true. They typically make a stronger claim: they say that TRANSITIVITY is a truism. A truism is a fundamentatum for further theorizing; they are things we theorize from, not theorize to. But TRANSPARENCY is also arguably a truism. (Or, as I’ll show in § 3.2, if TRANSPARENCY is not a truism, a thesis near enough for our purposes is). So now we have a puzzle: two principles, both of which are central to the study of consciousness and highly compelling, are in tension with each other. I won’t take sides on which is true. But I do assume the best theory will, on balance, be a theory that has the resources to grant both claims. If I am right, there is no such theory.5

The final paper—“Self-Consciousness Despite Elusiveness”—concerns ELUSIVENESS, and to a lesser degree TRANSPARENCY.6 On a subjectivist theory of consciousness, self-consciousness is a constitutive feature of phenomenal consciousness. Subjectivism cuts across the higher-order versus first-order theory divide. While all higher-order theories entail some form of subjectivism,

---

5 Well, there is no such theory assuming that one does not adopt a deflationary or dispositional understanding of the type of awareness at play in TRANSITIVITY. See § 3.3 for more.
6 The connections with TRANSPARENCY are subtle, so I won’t go into any meaningful detail here. But to foreshadow briefly, the connection concerns the positive component of TRANSPARENCY. The positive component may be understood as speaking to what we are in fact aware of when undergoing a conscious experience. This is fairly standard, and how we pitched things above. But the positive component can also be understood as speaking to the sense in which experience is ontologically reticent, i.e. the extent to which what we are aware of is appreciated as falling under certain ontological kinds (e.g. as being mind-independent). The claim I advance, following Frey (2013), is that when properly understood, this aspect of TRANSPARENCY’s positive dimension may have some interesting implications for a basic form of self-conscious. For more on these issues, see § 4.5.
first-order forms of subjectivism are viable as well, and are especially popular in the Heidelberg school and phenomenological traditions. But all subjectivist theories face a basic challenge: **ELUSIVENESS.** If **ELUSIVENESS** is true, isn’t subjectivism false?

I argue that the answer to this question is *no*. Part of the trouble is that **ELUSIVENESS**, as noted, is often misunderstood. So the first step (§ 4.2) is to get a clearer sense of what **ELUSIVENESS** amounts to. Having done this, I’ll argue that the truth of **ELUSIVENESS** actually places a non-trivial constraint on subjectivism, viz. that the self is not manifest as an object of consciousness (i.e. that self-consciousness is not a species of ‘object-consciousness’). Taking a cue from the Jesuit philosopher Bernard Lonergan, my aim in the second half of the paper (§ 4.4 – 4.5) is to lay the foundation for a version of subjectivism that meets this constraint. The resultant theory—what I call ‘Egocentric Representationalism’—is a significant advance over its competition (e.g. adverbialism about self-consciousness), and thus the best going for someone who finds both subjectivism and **ELUSIVENESS** compelling.

Each paper in the dissertation attempts to answer a different type of question. “Verbal Disputes in the Theory of Consciousness” asks a second-order question. It asks, not whether **TRANSLATIONALITY** is true, but why various philosophers disagree about **TRANSLATIONALITY**. “Transitivity and Transparency” asks a first-order question about **TRANSLATIONALITY**, viz. whether it is compatible with **TRANSPARENCY**. Finally, “Self-Consciousness Despite Elusiveness” does not attempt to argue for

---

7 The “Heidelberg School” is named after the German philosopher Dieter Henrich’s seminars in Heidelberg. The approach of the Heidelberg School is broad, drawing on resources from German idealism, neo-Kantianism, and the phenomenological traditions. For a nice representation of the Heidelberg School and its treatment of self-consciousness, see Zahavi (1999, 2007) and Frank (2007).
subjectivism. Nor do I argue for ELUSIVENESS. Rather, I am concerned with articulating what the subjectivist ought to say if she accepts ELUSIVENESS. In other words, it is advice for subjectivists. \(^8\)

\(^8\) That’s not to say that Egocentric Representationalism will have no implications for the viability of the subjectivist program more generally. Clearly, getting a grip on how we can be self-conscious despite there being a robust sense in which the self is not phenomenally manifest will remove a major obstacle to subjectivism. Of course, problems will remain; ELUSIVENESS is not the sole motivation for anti-subjectivism. For instance, experimental and clinical findings like somatoparaphrenia and thought insertion may present their own difficulties. For more on these latter issues, see e.g., Campbell (1999), Lane (2015), and Billon & Kriegel (2015).
2 VERBAL DISPUTES IN THE THEORY OF CONSCIOUSNESS

If we agree...not to worry about which mental phenomena deserve the honorific title 'consciousness', it may seem that there is nothing left about which Block and I disagree. We might even agree to apply the term 'conscious', in a special sense, to states that exhibit only thin phenomenality.

- David Rosenthal (2002a: 660)

2.1 THE TRANSITIVITY PRINCIPLE

The relationship between what makes a mental state conscious—i.e. state-consciousness—and our being aware of our mental states is a contested one. Consider:

(1) Carmichael's desire for tiramisu is conscious.
(2) Carmichael is aware of his desire for tiramisu.

According to Higher-Order (HO) theories of consciousness, if (1) is true (2) is true. HO theorists codify (2)'s entailment by (1) as 'The Transitivity Principle', or TRANSITIVITY for short:

TRANSITIVITY: A mental state of which one is wholly unaware is not conscious.

TRANSITIVITY is the theoretical backbone of HO approaches to consciousness. There are different ways of glossing the ‘unaware’ (or ‘aware’) in TRANSITIVITY, and these ways are what—at least in

---

9 This is a claim about what state-consciousness consists in: that Carmichael's desire for tiramisu is conscious consists in Carmichael’s being (suitably) aware of it. It is not a claim about what causes a mental state to be conscious, and nor is it a purely extensional thesis (cf. Kriegel 2009: 143).

10 Three brief notes. First, I will switch between this formulation and its contrapositive (‘conscious mental states are mental states we are aware of in some way’) throughout the dissertation for reasons of convenience. Second, HO theorists will sometimes use ‘conscious’ as well as ‘aware’ in TRANSITIVITY (e.g. Rosenthal 2005: 4), and I will follow suit in the present chapter and elsewhere. Third, by ‘aware’ (and ‘conscious’) I am concerned only with its non-deflationary, occurrent sense (§ 3.3). So, while in § 3.3 I will briefly look at deflationary and dispositional readings of TRANSITIVITY, for now I exclude those variants from my analysis.
part—individuate distinct HO theories. For instance, *Higher-Order Thought* theorists (e.g. Rosenthal 2005) say that Carmichael is conscious of his desire by having an appropriate higher-order thought (a ‘HOT’) about his desire; *Higher-Order Perception* theorists (e.g. Lycan 1996) will say that Carmichael is conscious of his desire by having an appropriate higher-order perception (a ‘HOP’) about his desire. But bracketing these differences, if *TRANSITIVITY* is false *simpliciter*, then HO theory is false *simpliciter*.  

Proponents of *First-Order* (FO) theories of consciousness deny that (1) entails (2), and so deny *TRANSITIVITY*. FO theorists are a motley crew. One can be a naïve realist (e.g. Campbell 2002) or a qualia realist (e.g. Shoemaker 1994; Block 1996) or a representationalist (e.g. Dretske 1995; Tye 1995; Bourget 2010) and be a FO theorist. FO theories are thus best defined negatively: a theory T is a first-order theory of consciousness if, according to T, awareness of a mental state is not necessary for that mental state’s being conscious. If the first-order approach is true, *TRANSITIVITY* is false. Conscious states may be states we are conscious *with*, but they are not necessarily states we are conscious *of* (Dretske 1993: 281).  

11 Some HO theorists (e.g. Berger 2014; Brown 2015) advance instead of (2) the slightly different (2*):

2.* Carmichael is aware of himself as desiring tiramisu.

Then *TRANSITIVITY* would really be the thesis that (1) entails (2*). Call this version *TRANSITIVITY**. (It is also sometimes said that if by mental states we mean *token* mental states, *TRANSITIVITY* actually entails *TRANSITIVITY*—see, e.g. Rosenthal 1997a; Kriegel 2009). Although we will discuss *TRANSITIVITY* later (cf. § 3.4.3, fn. 68), since nothing concerning the argument from the present chapter turns on employing *TRANSITIVITY* *in lieu of* *TRANSITIVITY*, I will stick with the latter going forward.

12 I assume FO and HO theorists are targeting *phenomenal* consciousness when they are targeting state-consciousness—i.e. the property a mental state has when it is such that it is like something for its subject. That said, it is not obvious that *all* HO theorists see themselves as what Block (2011a) calls “ambitious” HO theorists in that they purport to be capturing consciousness as understood by Nagel’s (1974) ‘what it is like’-locution. Folks like Rosenthal, Weisberg, and Kriegel are clearly ambitious in this sense, but less clear is Lycan. Non-ambitious HO theorists are excluded from this chapter’s argument.

13 In saying that FO theorists reject *TRANSITIVITY*, I do not mean to claim that FO theorists will claim that we *necessarily* cannot be aware of our conscious states, or even that we are unaware of most of our conscious
In our epigraph, David Rosenthal—arguably the most prominent HO theorist—suggests that his disagreement with Ned Block (and presumably other FO theorists) comes down to a disagreement about what phenomenon deserves the honorific ‘consciousness.’ The disagreement Rosenthal is referring to is the disagreement over whether TRANSITIVITY is true. Now Rosenthal is almost certainly being facetious. But I’ll argue that his speculation, made in jest or not, is more right than he likely realizes: the dispute over TRANSITIVITY, I’ll claim, is verbal.

By ‘verbal’ I do not mean non-substantive. Many disputes that are verbal are also non-substantive. Indeed, the paradigmatic examples—like disputes concerning whether glasses are cups—are plausibly like this. Nothing turns on these disputes. However, one can have a verbal dispute that is still substantive. And that is precisely what is going on here: the dispute over TRANSITIVITY, though verbal, is nevertheless substantive. The upshot is a unique methodological recommendation. Normally when a dispute is verbal and non-substantive, the methodological recommendation is to stop talking about the topic of that dispute (Jenkins 2014). This is not the appropriate recommendation for a dispute that is verbal and substantive. What that appropriate recommendation is—and how it shakes out in the particular case of TRANSITIVITY—will be explored at the end of the chapter (§ 2.4).

2.2 USING AND ABUSING THE NAGELIAN CONCEPTION OF CONSCIOUSNESS

We begin with the Nagelian Conception of state-consciousness—‘Nagelian’ for short. Inspired by Thomas Nagel (1974), it can be framed like so:

states. (Although Tye, at least insofar as FO theorists go, might be an exception. See § 3). Relatedly, accepting TRANSITIVITY is consistent with saying—as many HO theorists do—that we can also be aware of something else in virtue of being in a conscious state.

VERBAL DISPUTES IN THE THEORY OF CONSCIOUSNESS
**Nagelian**: For any mental state $M$ of a subject $S$, $M$ is conscious iff $M$ is such that there is something it is like for $S$ to be in $M$.\(^{14}\)\(^{15}\)

There is something curious about the role Nagelian plays in discussions concerning the nature of consciousness. Nearly all FO and HO theorists agree that Nagelian is true.\(^{16}\) Talk of ‘what it is like’ is common currency. Yet many HO theorists are keen to point out that not only does a commitment to Nagelian entail a commitment to Transitivity, but also that to think otherwise is somewhat shocking—that is, if one thinks otherwise, one is abusing or at the very least misunderstanding Nagelian. Here are two examples.

The first example concerns an exchange between Block (2011a) and the HO theorist Josh Weisberg (2011). Recapitulating Block’s criticism of HO theories, Weisberg tells us:

---

\(^{14}\) With Charles Siewert, ‘something it is like for $S$ to be in $M$’ is used in a non-derivative sense: there isn’t something it is like for $S$ to be in $M$ only insofar as some other state $M^*$ accompanies $M$, and there is something it is like to be in $M^*$ (2011: 244-45). This does not prejudice the matter against relational HO theories: that consciousness is a non-derivative property doesn’t mean consciousness is a non-relational property. It is consistent with saying that a mental state $M$ is non-derivatively conscious with saying that $M$ is conscious only when it stands in a relation to some other mental state $M^*$. What *would* be inconsistent would be adding the further claim that $M$ is conscious only because $M^*$ is itself conscious.

\(^{15}\) The phrase ‘$M$ is such that’ is included in my formulation of Nagelian to help bring out Siewert’s point (see fn. 14). However, there is a slight risk of misunderstanding here, for one might read the phrase ‘$M$ is such that’ to suggest that what is relevant to whether $M$ is conscious is how $M$ *itself* is, or is intrinsically. This way of reading Nagelian might load the dice against some HO theorists (e.g. Rosenthal), for on these variants of HO theory it is not $M$ itself that matters. Fortunately, the ‘$M$ is such that’ phrase does not need to imply that $M$ undergoes some intrinsic modification when it is conscious. *That* is certainly a thesis that many HO theorists would reject. The ‘$M$ is such that’ phrase is compatible with $M$ simply being the intentional object of a state of higher-order awareness. Here, $M$ would be such that it is like something for $S$ iff $M$ is such that it is the intentional object of a state of higher-order awareness. Another option is to use the following alternative formulation:

**Nagelian**: For any mental state $M$ of a subject $S$, $M$ is conscious iff there is something it is like for $S$ to be in $M$.

The advantage of Nagelian* is that it shifts any (even if only apparent) emphasis from what a mental state is like to what its like to be in a mental state. Since nothing in what follows will turn on our employing Nagelian* in lieu of Nagelian, I’ll stick with the latter.

\(^{16}\) For more agreement on Nagelian from both sides of the aisle, see, e.g.: Block (1995: 228), Dretske (1999: 103), Tye (1995: 5, 133), Kriegel (2009), Rosenthal (1997b), and Hellie (2007a, 2007b), and § 1.

VERBAL DISPUTES IN THE THEORY OF CONSCIOUSNESS
[According to Block], [m]y defense entails that the what-it’s-like-ness occurring during an episode of HO misrepresentation ‘is fake’! There is no first-order state, and so...there is no property of what-it’s-like-ness...And by this logic, the HO theorists cannot explain why there is an important difference between those really having conscious pains (replete with real what-it’s-like-ness) and those merely representing themselves as having conscious pains (possessing only fake what-it’s-like-ness). Block concludes that there is no way to explain why pain matters on the HO approach. And that means that the approach is ‘about consciousness in a merely technical sense of the term.’ This is not meant as a compliment. (2011: 439, italics mine)

Yet Weisberg is not convinced:

According to Block’s reading, pains matter even if the subject is in no way aware of them. Consider a person in such a state. She will not be aware of being in pain. She will deny that anything is wrong...But according to Block, the pains that matter are still present. The subjects might even be in agonizing pain—that is, they might possess states with the what-it’s-like-ness property of agonizing pain—even though there’s nothing it’s like for the subjects. I contend that, whatever is going on here, to characterize such episodes as conscious is to use a ‘merely technical sense of the term’. (2011: 440, italics mine).

Weisberg gives Block right back what Block gave him. To characterize mental states of which a subject is wholly unaware as conscious—to reject TRANSITIVITY—is to employ a “merely technical sense of the term.” Block didn’t mean this as a complement. Weisberg, I take it, didn’t either.17

A second example concerns Block’s stance on visual extinction. Visual extinction is caused by damage to the parietal lobe. Extinction subjects can report stimuli on either side of their visual field so long as the stimuli are not presented simultaneously. However, if both stimuli are presented at once—one on the unimpaired side of the visual field and one on the impaired side—such subjects can only report the stimulus on the unimpaired side. In this sense, perception of stimulus on the unimpaired side ‘extinguishes’ perception on the impaired side. Yet Block allows

---

17 The backstory concerns a version of the ‘radical misrepresentation’ objection to HO theories, which Block deems fatal. Nothing turns on the details, but for more on this objection see fn. 69.
that such subjects can “have phenomenal experience of those stimuli without knowing it” (2001: 203). Rosenthal is dubious, to say the least. He says:

Block...allows that phenomenality can occur not only without one’s knowing it, but in cases in which one would firmly deny its occurrence. This does not fit comfortably, however, with the explanation of phenomenality as “what it is like to have an experience.” It’s important to distinguish this somewhat special use of the phrase what it’s like to describe subjectivity from its more general, non-mental use. There is something it’s like to be a table, or even to be this very table. What it’s like to be a table...is roughly something’s having characteristic features of tables.

But this is of course not what’s involved in talking about what it’s like to have an experience. As Nagel stressed...what it’s like to have an experience is what it’s like for the individual that has the experience. When a person enjoys the taste of wine...there is something it’s like for that person...

Not so in cases of visual extinction; there is nothing it’s like for an extinction subject to have a qualitative experience of the extinguished stimuli. That’s why seeing visual extinction as the having of phenomenality without one’s knowing it does not fit comfortably with the explanation of phenomenality in terms of what it’s like to have an experience. (2002: 656)

Rosenthal’s response to Block mirrors Weisberg’s response in an important respect. Weisberg is not merely saying that Block is wrong to characterize states of which a subject is wholly unaware as conscious, but in some sense obviously wrong—hence his charge that Block “abuses” Nagelian. Similarly, Rosenthal suggesting there is something that answers to the question ‘what it is like to see the extinguished stimuli’ only in the way there is something that answers to the question ‘what it is like to be a table’. That’s a strong claim.

In sum, the presumption—at least on the part of the HO theorist—seems to be this: if one thinks transitivity is false, one cannot have Nagelian in mind, protestations aside. This seems to be the real import of Weisberg’s ‘right-back-at-you’ to Block. Weisberg is not just being flippant.

---

18 This point is key, since it shows that this is not just about reportability. Block confirms this elsewhere too, saying that extinction cases show that “a subject could have an experience that he does not and cannot know about” (2008: 289). Fred Dretske—another FO theorist—concurs (see, e.g. 2006: 167).
Quite the contrary, he is insisting that Block must be talking about something else entirely, given NAGELIAN. These observations raise the following possibility in an acute form: could HO and FO theorists disagree about TRANSITIVITY because they disagree, perhaps implicitly, on what they mean by NAGELIAN? I think we can make a strong case that the answer is yes.

2.3 THE TRANSITIVITY DISPUTE IS A VERBAL DISPUTE

2.3.1 PRELIMINARIES

Let a dispute be any exchange that involves (at least) two parties disagreeing over a sentence S, where one assertively utters, or otherwise endorses S, and the other assertively utters or otherwise endorses ¬S (Jackson 2014). Finding conditions that are jointly necessary and sufficient for a verbal dispute is not easy. Luckily, a sufficient condition will do. Here I follow Eli Hirsch. According to Hirsch, two parties A and B are in a verbal dispute over S if A and B are engaged in a sincere prima facie dispute, but, given the correct view of linguistic interpretation, A and B will agree that each “speaks the truth in its own language” (2009: 239).19 We’ll have cause to examine Hirsch’s diagnostics in detail below, but I’ll note now that our reliance on Hirsch’s account of ‘the correct view of linguistic interpretation’ in the argument to follow will be minimal. This is important, for it blunts the extent to which the argument to come stands and falls with the particularities of Hirsch’s view.

---

19 More accurately, it’s an imagined public language. For Hirsch, Tyler Burge’s semantic externalism commits us to this stipulation (1979). Matthew McGrath nicely points out that Hirsch’s move, in effect, is to reinterpret the phrase “in one’s own language” such that “speaking the truth in one’s own language isn’t the same as speaking truly using the language one speaks” (2008: 485). For doubts about the necessity of this stipulation, see Jackson (2013).
We saw that all parties agree on Nagelian. So, employing Nagelian, Transitivity can be translated as:

**Nagelian Transitivity:** A mental state of which one is wholly unaware cannot be such that it is like something for one to be in it.

This allows us to frame our main thesis with some more precision. I’ll argue that:

HO and FO theorists disagree about the truth-value of Transitivity because HO and FO theorists are having a (tacit) semantic disagreement about the consequent of Nagelian Transitivity, viz. about the meaning of the expression ‘it is like something for one’; and (ii) HO and FO theorists would agree about the truth-value of Transitivity if they agreed about the meaning of ‘it is like something for one’.

This claim requires two refinements. First, even if their dispute is verbal, HO and FO theorists may not be faultless in every respect (Eklund 2009: 143). There may be facts about the correct use of the expression ‘it is like something for one’, and one side might just get things wrong. Take the following case, described by Karen Bennett:

Consider...a dispute you might have with someone who insists upon using the English word ‘telephone’ to refer to leprechauns. Suppose that the two of you agree that the world contains certain sorts of communication devices, and does not contain little green people who hide gold at the end of rainbows. You say that there are telephones; he says that there are no telephones. Although this is paradigmatically a verbal dispute, you win. Facts about the correct use of the English expressions in the sentence, conjoined with facts about what sorts of entities we are presuming the world to contain, dictate that ‘there are telephones’ is determinately true (2009: 40).

Thus I’ll assume a verbal dispute can arise due to one side misusing a portion of language that the other side is using correctly. This point will be important as we proceed.

Second, and relatedly, I do not claim that the dispute over Transitivity is non-substantive. I briefly remarked on this possibility earlier. To say that a dispute is verbal but still substantive is
one way of saying that a dispute is not *merely* verbal.\(^{20}\) Suppose we have a disputed sentence \(s\), where A asserts \(s\) and B asserts \(\neg s\). For A and B’s dispute over \(s\) to be a non-substantive verbal dispute, their dispute must be (of course) verbal, but the rival interpretations \(L\) and \(L^*\) on which A and B can agree that \(s\) is true must also be in *metaphysical parity*. There are two primary options for cashing out the notion of metaphysical parity.\(^{21}\) The first is *intensional equivalence*: \(L\) and \(L^*\) are metaphysically on par if they are intensionally equivalent (e.g. Hirsch 2011).\(^{22}\) Consider the following sentence disputed by mereological nihilists and anti-nihilists:

\[(T) \quad \text{There are tables.}\]

On Hirsch’s view, there is an interpretation of \((T)\) according to which \((T)\) is true if and only if there are simples arranged table-wise. Granting that everyone believes in simples, the anti-nihilist and the nihilist agree that, on this interpretation, \((T)\) is true. But because this interpretation is intensionally equivalent to the other available interpretations—‘There are tables’ expresses the same intension as ‘There are simples arranged tablewise’—the dispute over \((T)\) is not just verbal, but non-substantive.

---

\(^{20}\) There are other ways to gloss the distinction between verbal and merely verbal disputes (e.g. Chalmers 2011; Jenkins 2014). For example, Chalmers points out that sometimes words matter: “Questions about what falls into the extension of ‘marriage’ and ‘murder’ may in some sense be verbal, but the answer to these questions may make a serious difference to people’s lives” (ibid: 518). So ‘marriage’ is unlike, say, ‘cup’. If mere verbalness is a matter of lack of practical import, I’m inclined to say that this is another way in which the dispute over \(\text{TRANSITIVITY}\) is not merely verbal.

\(^{21}\) These options are ‘primary’ in the sense that they are the most prevalent options in the verbal disputes literature. I will discuss one other model—the ‘transmission’ model—in § 2.4.

\(^{22}\) It is common to understand intensional equivalence in terms of *intensional resources*, i.e. possibility-expressing power. For instance, where a set of possible worlds provide the truth conditions for a sentence, Daniel Liebesman says: “two languages differ in intensional resources iff there is some set of possible worlds that provides the truth conditions for a sentence of one language, but no sentence in the other” (2015: 305). Deflationists like Hirsch will insist that proponents of reputedly different ontologies (like the nihilist or the universalist about physical objects) should be seen as simply using different linguistic frameworks to describe the same set of possible worlds.
The second is some rough combination of intensional equivalence and the relative ‘joint-carving-ness’ of \( L \) and \( L^* \) (e.g. Sider 2012). Consider the following example, borrowed from Jonathan Jacobs (forthcoming), by way of Sider (2012). Jacobs has us consider a rectangle formed by conjoining two squares, the left of which is white and the right of which is green.

Figure 2.3.1  *Naturalness and Gerrymandered Predicates*

It is true that the rectangle is half white and half green. Likewise, it is true that the area of the rectangle that is green is equal to the area of the rectangle that is white (*ibid*).

But then Jacobs has us consider a different linguistic community. This community does not have our color concepts. Rather than conceptualizing the rectangle as divvied in half from top to bottom, they conceptualize it as divvied color-wise, split by the dotted line running across the figure on a diagonal (*ibid*). Their concept for the color of the top left triangle is `wheen', and their concept for the color of the lower right triangle is `grite'.

Now suppose a member of this community—call him Carmichael—said, “The area of grite is equal to the area of wheen". Does Carmichael speak truly? Jacob points out that given the concepts he employs, the answer is yes. After all, the area of the grite isn’t smaller than the area of wheen, and it isn’t larger (Jacobs, *ibid*). Nevertheless, as Sider puts it, “it is nearly irresistible to describe these people as making a mistake" (2012: 2). Even though they may assert truths using their color concepts, they are “missing something" about the objective structure of reality (*ibid*). One can speak truly but do so with the wrong concepts. So the mistake is a conceptual one.
The propositions Carmichael expresses by his concepts fall short of getting at the fundamental structure of reality. Unlike ‘white’ and ‘green’, ‘wheen’ and ‘grite’ (or ‘grue’ and ‘bleen’) are unnatural. They are gerrymandered. They don’t, as Plato put it, carve nature at the joints. For Sider, the epistemic norms of inquiry don’t stop at truth.

These observations from Hirsch and Sider generalize, and so can be used as a framework for understanding the dispute over TRANSITIVITY. I contend that our alternative languages—the languages HO and FO theorists are charitably interpreted as speaking—will not be in metaphysical parity. This incongruity in metaphysical parity won’t be due to an incongruity in the joint-carving status of the relevant terms in these languages—or if it is, this is not an issue I’ll broach here. For Sider, structure is a primitive theoretical posit, and his epistemology of structure is thoroughly Quinean. But this won’t matter. For there will be a clear failure of parity due to lack of intensional equivalence. Thus, I’ll claim that the dispute over TRANSITIVITY is a substantive verbal dispute. I discuss this result further, along with the more general implications it carries for the dispute between HO and FO theories, in § 2.4.

So much for refinements. My argument unfolds in three steps. The first step (§ 2.3.2) appeals to Daniel Stoljar’s (forthcoming) recent work on the semantics of ‘what it is like’-sentences. Examples of ‘what it is like’-sentences include:

(a) There is something it is like to smell garlic.
(b) There is something it is like for Carmichael to see the sunset.
(c) What is it like to hear Led Zeppelin?

NAGELIAN TRANSITIVITY is unlike (a) and (b) as it does not include an explicit quantifier, and unlike

---

23 I don’t adjudicate the correct semantics for ‘what it is like’-sentences in the argument to follow. I’m also ignoring views on which NAGELIAN is meaningless (Hacker 2002), false, or trivial (Snowdon 2010).
(c) as it is an indicative instead of an interrogative. Nagelian transitivity is also a less common English sentence than (a), (b) and (c). All the same Nagelian transitivity is a ‘what it is like’-sentence, even if, like (a) and (b) but unlike (c), it does not explicitly use the term ‘what’.

The second and third steps (§ 2.3.3 - § 2.3.4) build on the second, employing Stoljar’s different semantics for ‘what it is like’-sentences to build plausible alternate languages for the FO and HO camps. These languages will yield distinct truth-conditions with respect to Nagelian transitivity (and by extension, transitivity). I’ll then argue that FO theorists are best interpreted as belonging to a linguistic community in which Nagelian transitivity (and transitivity) is false, and HO theorists are best interpreted as belonging to a linguistic community in which Nagelian transitivity (and transitivity) is true (§ 2.3.3). I follow Hirsch in assuming that principles of charity are key to the norms of linguistic interpretation. Hirsch focuses on three such principles, but we will for the most part only make use of charity to understanding: ceteris paribus, we ought to assume that the speakers of a language don’t make a priori or conceptual errors concerning basic matters of fact (2011: 182).24 The final step (§ 2.3.4 - § 2.3.5) claims that we can extract from our charitably assigned languages a pair of undisputed sentences, one that is

24 The other two are charity to perception and charity to retraction. Though charity to retraction will come up a bit (see fn. 35), charity to perception—i.e. the claim that interpretation shouldn’t result in attributing pervasive basic perceptual errors to speakers—will not. Charity to perception is apt when the dispute concerns entities that are actually publicly available to perceptual awareness (e.g. disputes concerning ordinary physical object ontology). This condition is plausibly not met in the dispute over transitivity, as it concerns conscious mental states.
false and analytically equivalent to Nagelian Transitivity according to FO theorists, and one that is true and analytically equivalent to Nagelian Transitivity according to HO theorists.25

2.3.2 The Semantics of ‘What it is Like’-Sentences

Stoljar (forthcoming) offers five accounts of the semantics of ‘what it is like’-sentences, yet we’ll only make us of two.26 Also, our presentation will be somewhat simplified. It is the way these

25 One might worry that the meaning of ‘what it is like’-sentences is a technical matter, making the semantic project futile. On this sort of view, ‘what it is like’-sentences involve terms that may look like colloquial terms but in fact have a distinct meaning determined—as Stoljar puts it—by the “technocrats” (i.e. philosophers of mind). Call this view the Technical Account (e.g. Lewis 1995: 140). Talk involving ‘what it is like’ would be like talk about ‘work’ in physics. I think the Technical Account is false (Stoljar, forthcoming; Farrell, forthcoming). Following Jonathan Farrell (ibid), the Technical Account of ‘what it is like’-sentences can be defined as the conjunction of the following three theses:

Technical: ‘What it is like’-sentences involve technical terms.
Introduction: Philosophers introduced these technical terms.
Meaning: It is because of the special meaning that these terms have that we can use ‘what it is like’-sentences to talk about consciousness.

Farrell argues forcefully that both Technical and Introduction are false. Briefly on Technical: Farrell points out there is no difference in meaning between the non-philosopher’s use of ‘what it is like’-talk and the philosophical use. But if the alleged technical terms in ‘what it is like’-talk have the same meaning as their “everyday, non-technical, look- and sound-alike counterparts,” then they are not technical at all. Hence, Farrell contends that Technical is false. Briefly on Introduction: Farrell rightly notes that technical terms are typically explicitly introduced as such. This is because the articulation of a technical term sounds and looks just like its ordinary counterpart. So if the technical account were true, we would expect Introduction. But, per Farrell, that’s not what we find. And he is right: Nagel doesn’t do this, and neither do Wittgenstein (1980), Farrell (1950), or Sprigge (1971) before him.

26 A few notes. First, Stoljar is concerned with ‘what it is like’-sentences in their stereotypical context of use. By ‘stereotypical’ Stoljar means the ways of using such sentences that speakers regard as routine though even if not mandated by their linguistic meaning. Stoljar uses the example ‘he has all the qualities of his father’ to convey this idea (ibid). If we only consider this sentence’s linguistic meaning, the sentence could express propositions about any qualities. But it is good qualities that are relevant in its stereotypical. Second, Stoljar is concerned with explicating the linguistic meaning of ‘what it is like’-sentences and the propositions—here, truth-conditions—expressed in a context by such sentences, not those that are, say, conveyed (ibid). Stoljar also assumes ‘what it is like’-sentences are routinely used by speakers to convey propositions about experiences and subjects of experiences. I follow him on both of these scores. Further, I’ll also assume that whenever speakers use an expression, they do so with beliefs about it’s meaning, be they tacit or explicit.

VERBAL DISPUTES IN THE THEORY OF CONSCIOUSNESS
views differ regarding the prepositional ‘for the subject’-phrase that is part of Nagelian and part of (even if only implicitly) all ‘what it is like’-sentences that will be critical for our purposes.

We begin with a single canonical ‘what it is like’-sentence. Let Canonical be the sentence formed by swapping ‘see the sunset’ for a general perceptual act-object schema ‘ϕ[‘ing] x’ in ‘There is something it is like for Carmichael to see the sunset.’ This gives us: ‘There is something it was like for Carmichael to ϕ x.’ Then, according to the Affective View:

The Affective View: Canonical is true iff there is some way that Carmichael feels as a result of Carmichael’s ϕ’ing x.

The Affective View is Stoljar’s preferred story. On this picture, ‘what it is like’-sentences assert “an experiential relation between an event and an individual, where an experiential relation obtains just in case there is a way the individual feels in virtue of the event” (ibid).

One might wonder what it means to say that an individual ‘feels’ some way on the Affective View. If by ‘feels’ we mean something like a bodily sensation, then the Affective View is likely false. Suppose we consider the following sentence: ‘There is something it is like for Mary to see red.’ Stoljar points out that if the Mary in question is Frank Jackson’s Mary—the physically omniscient neuroscientist who from birth is locked in a black-and-white room—this sentence reports an experience (ibid). But then, though the sentence is true, Mary needn’t feel any way in virtue of having her experience. It’s not as if she must have touched something, or have an emotion, for seeing red to be like something for her (ibid). Likewise for Canonical: Carmichael need not feel any way as a result of seeing the sunset.

Stoljar offers two responses to this worry, each of which can be employed by a friend of the Affective View. His first response acknowledges that while some philosophers have employed a
narrow notion of feeling, English speakers are more lax. Consider the notion of feeling at play in ‘I feel like class has taken forever’. This credibly reports an experience, but the sense of feeling involves no appeal to bodily sensations or emotions.\(^{27}\)

The second response requires a modification of the Affective View. In effect, it treats the Affective View as a disjunction: CANONICAL is true iff (i) there is some way that Carmichael feels as a result of Carmichael’s \(\varphi\)’ing \(x\) or (ii) there is some way that \(x\) seems to Carmichael in virtue of his \(\varphi\)’ing \(x\). This avoids the objection from Mary, since it is plausible that things seem some way when she sees something red, even though she doesn’t feel any way (\textit{ibid}).

Going forward, I will largely operate under the first response, where we restrict attention to feeling in either the broad or narrow sense. However, when a move to the ‘disjunctive’ version of the Affective View might make a difference, this will be noted.

On the Affective View, the ‘for the subject’-phrase is (i) logically speaking an argument of the verb phrase ‘is like’, and (ii) semantically expressing affectiveness, i.e. some way the subject feels as a result of \(\varphi\)’ing. Note what this doesn’t imply. It does not follow that when Carmichael sees the sunset he is aware of his event of seeing the sunset. It only means that Carmichael feels a certain way \textit{in virtue} of seeing the sunset (\textit{ibid}). This point will prove crucial below.

The second account is the \textit{Operator View}:

\textbf{The Operator View:} CANONICAL is true iff there is some way such that it seems to Carmichael that Carmichael’s \(\varphi\)’ing \(x\) is that way.

The Operator View is so-called because it appeals to an operator—a function from sentences to sentences—viz., ‘it seems to you that.’ This approach agrees with the Affective View on the

\(^{27}\) Brogaard (2012) does a particularly nice job of canvassing the broader notions of feeling.
structure of CANONICAL, but departs over the ‘for Carmichael’ in CANONICAL. On the Operator View, the ‘for Carmichael’ phrase is (i) logically speaking an operator, and (ii) semantically expresses what ‘it seems to you that’ expresses (ibid). This too will prove critical as we proceed.

There are, as Stoljar notes, tough questions about how ‘seems’ is to be interpreted. ‘Seems’ will at the very least implicate awareness—if x seems F to Carmichael then Carmichael is aware of x as being F—but ‘awareness’ itself requires disambiguation. Eric Lormand (2004), for instance, takes ‘seems’ to implicate an inner sense view, akin to Lycan’s HOP theory. But whether ‘seems’ is best understood as a phenomenal, perceptual, or doxastic notion can be put aside for now. The present focus is this: just as Hirsch says that the culprit in the dispute over ontology is ‘there is’, the culprit in the dispute over TRANSITIVITY is ‘for the subject.’

2.3.3 Higher Order English and First Order English

We can now define two languages: ‘HO-English’ and ‘FO-English.’ According to HO-English, a ‘what it is like’-sentence (including NAGELIAN TRANSITIVITY) has its truth-conditions in accordance

28 Stoljar also mentions the Property View, according to which:

The Property View: CANONICAL is true iff there some property F such that the event of Carmichael’s φ’ing x has F.

The Property View option won’t figure into our discussion below, but it is worth mentioning as it treats the ‘for’ as a complementizer of the infinitival clause ‘Carmichael to φ’x’. Alex Byrne (2004) notes that this permits us to rewrite CANONICAL as:

CANONICAL*: For Carmichael to see the sunset is like something (i.e. has some property).

Which in turn is equivalent to:

CANONICAL**: For a sunset to be seen by Carmichael is like something.

If CANONICAL* and CANONICAL** are in fact equivalent to CANONICAL, ‘for’ will have no particular attachment to ‘Carmichael’. (Unlike, as Byrne points out, the ‘for’ in ‘The police are looking for Carmichael’). The ‘for Carmichael’-phrase is thus context-sensitive: some will require it, some won’t. This marks a significant structural difference from both the Affective and Operator Views.

VERBAL DISPUTES IN THE THEORY OF CONSCIOUSNESS
with the Operator View. According to FO-English, a ‘what it is like’-sentence (including NAGELIAN TRANSITIVITY) has its truth-conditions in accordance with the Affective View. The questions before us now are two-fold. First, ought the FO theorist interpret the HO theorist as speaking HO-English? Second, ought the HO theorist interpret the FO theorist as speaking FO-English? The answer to both questions is yes.

We can begin by adopting the standpoint of a FO theorist. If we are FO theorists trying to make sense of what HO theorists are saying when they assert TRANSITIVITY, things are fairly easy. Rosenthal—again, arguably the progenitor of modern-day HO theory—all but outright admits the Operator View. This is what Rosenthal says with respect to NAGELIAN: “As many, myself included, use that phrase, there being something it’s like for one to be in a state is simply its seeming subjectively that one is in that state” (2011: 434).

We also have implicit commitments to the Operator View. The literature is riddled with HO theorists stressing the complementizer ‘for’ in ‘what it is like for’—often literally, by underscoring or italicizing—as if to remind us that the term ‘for’ alone is supposed to tell us TRANSITIVITY must be true. Stoljar calls these, aptly, emphatic arguments. When Rosenthal expressed his bewilderment that Block could possibly entertain states of visual extinction being conscious (§2.2), he points to the ‘for’ in NAGELIAN. Similarly, Weisberg tells us that the ‘for the organism [i.e. our ‘for the subject’]’ in Nagel’s ‘something it is like for the organism’ suggests “a connection to the rest of the mind, a mode of access by a sentient subject” (2011: 411). Joseph Levine makes the same point:

---

29 Hellie (2007a), Weisberg (2011), Rosenthal (2011), and Janzen (2011) express explicit sympathies for some version of the Operator View. Everyone who has committed to this view in print is a HO theorist.
...the very phrase that serves to canonically express the notion of the phenomenal—“what it’s like for x to...”—explicitly refers to the phenomenal state in question being “for” the subject...Phenomenal states/properties are not merely instantiated in the subject, but are experienced by the subject. Experience is more than mere instantiation, and part of what that “more” involves is some kind of access (2007: 514).

I suggest that Rosenthal, Weisberg, and Levine are giving emphatic arguments here. Further, I suggest that emphatic arguments suggest an implicit commitment to the Operator View.

To see this, note how the emphatic argument unfolds. First, we begin the premise that NAGELIAN is true, with an added emphasis on the prepositional ‘for the subject’

\[
P_1 \quad \text{For any mental state } M \text{ of a subject } S, M \text{ is (phenomenally) conscious iff there is something it is like for } S \text{ to be in } M.
\]

And then from \( P_1 \) we move to the claim that some form of HO theory is true:

\[
P_2 \quad \text{There is something it is like for } S \text{ to be in } M \text{ only if } S \text{ is aware of } M.
\]

\[
\therefore \quad \text{HO} \quad \text{For any mental state } M \text{ of a subject } S, M \text{ is (phenomenally) conscious only if } S \text{ is aware of } M.
\]

Yet on the supposition that the emphasizing of the expression ‘for the subject’ itself counts for something, the move from \( P_1 \) (NAGELIAN) to \( P_2 \) and HO will likely only make sense if we presume a thesis concerning what that expression means, for instance:

\[
P_3 \quad \text{There is something it is like for } S \text{ to be in } M \text{ only if } M \text{ seems some way to } S.
\]

But \( P_3 \), of course, is just another way of putting the Operator View. Greg Janzen, another HO theorist, doesn’t pull any punches here: “the very language of the what-it-is-like formula, the

\[30\] Levine is not a HO theorist if we restrict HO theories to HO representational theories. But there is no reason to assume that the awareness-of relation we bear to our mental states is representational. Levine denies this (2006), as does Hellie on his ‘Higher-Order Acquaintance’ view (2007b).

\[31\] See Stoljar (forthcoming) for a similar point. If one thinks ‘seems’ is too strong, it’s worth noting that it is not essential to the Operator View. We have treated the prepositional ‘for the subject’ as semantically expresses whatever ‘it seems to you that’ expresses, because this is how (as noted) Rosenthal himself puts it. But we could just as well appeal to other sentential operators such as ‘Carmichael represents it as being
words in it, suggests that it ought to be read as expressing a proposition about a subject’s awareness of her own mental states” (2011: 283). So, as an interpretive matter, charity is not much of an issue when it comes to attributing HO-English to the HO theorist. To advance the emphatic argument, as so many HO theorists do, is to be committed to the Operator View. HO theorists, whether explicitly or implicitly, speak HO-English.

What emerges, then, is the following equivalence: in HO-English, NAGELIAN TRANSITIVITY (and by extension, TRANSITIVITY) is equivalent to ‘A mental state of which one is wholly unaware cannot be such that it seems some way to one.’ This is an equivalence, because in HO-English, a ‘what it is like’-sentence has its truth-conditions in accordance with the Operator View. Now, Bennett (2009: 53) has argued that for Hirsch’s condition to truly be sufficient, we cannot have any old equivalence. The equivalence in question must be an analytic equivalence. And of course she is right: surely there would be no verbal dispute if the parties didn’t disagree about the meaning of the disputed sentence. Bennett tells us that this is relevant for Hirsch’s claim that the dispute over physical-object ontology is verbal: because the relevant equivalences are not analytic in those cases, Hirsch is wrong. Here it seems fairly clear, however, that Rosenthal is intending to convey what he means by NAGELIAN. He says, “as many, myself included, use that phrase” (2011: 443, emphasis mine). To speak of a way one ‘uses’ a phrase is, typically, to speak of what a phrase means to one.

If the HO theorist is speaking HO-English, the FO theorist should agree that she speaks the truth in her own language when she asserts TRANSITIVITY and NAGELIAN TRANSITIVITY. We already saw

the case that’, or non-sentential operators like ‘Carmichael is aware of’ (Stoljar, forthcoming). What matters for present purposes is that ‘for the subject’ expresses awareness of a mental state.

VERBAL DISPUTES IN THE THEORY OF CONSCIOUSNESS
this a moment ago. On the Operator View, there is something it is like for Carmichael to \( \varphi \) \( x \) iff there is some way such that it seems to Carmichael that Carmichael’s \( \varphi \)’ing \( x \) is that way. From here, the argument to NAGELIAN TRANSITIVITY and TRANSITIVITY is swift. First, by the Operator View what seems some way to Carmichael is Carmichael’s state of \( \varphi \)’ing. Carmichael’s state of \( \varphi \)’ing is just his state of seeing the sunset, i.e. a mental state. Yet if Carmichael’s state of \( \varphi \)’ing is such that it is like something for Carmichael iff it seems some way to Carmichael, and Carmichael’s state of \( \varphi \)’ing is conscious iff it is such that it is like something for Carmichael, then Carmichael’s state of \( \varphi \)’ing is conscious iff Carmichael is in some way aware of \( \varphi \)’ing. For Carmichael’s state of \( \varphi \)’ing cannot seem some way to Carmichael without Carmichael being in some way aware of his state of \( \varphi \)’ing. So if the Operator View is true, then TRANSITIVITY and NAGELIAN TRANSITIVITY are true as well.

Let us now adopt the standpoint of the HO theorist: is it charitable to interpret the FO theorist as speaking FO-English? It seems that it is. For notice that, from the standpoint of the HO theorist, the FO theorist is making an a priori error when he denies NAGELIAN TRANSITIVITY. For given the Operator View, it follows a priori that NAGELIAN TRANSITIVITY is true. It is plausibly an a priori truth that if a state of \( \varphi \)’ing is like something for Carmichael iff that state of \( \varphi \)’ing is such that it seems some way to Carmichael, then a state of \( \varphi \)’ing is like something for Carmichael iff Carmichael is in some way aware of that state of \( \varphi \)’ing. This is so because it is plausibly an a priori truth that if a state of \( \varphi \)’ing is such that it seems some way to Carmichael, then Carmichael is in some way aware of that state of \( \varphi \)’ing.

And note that this is an exceedingly reasonable application of charity to understanding. As Jackson remarks, we cannot be required to avoid attributing any a priori mistakes at all to a

VERBAL DISPUTES IN THE THEORY OF CONSCIOUSNESS
subject (2013: 427). It is not hard to fathom how a rational subject might have trouble recognizing a complex a priori truth, like one stemming from a tough mathematical equation, say. However, the relevant a priori entailment, from ‘Carmichael’s ϕ’ing x seems some way to Carmichael’ to ‘Carmichael is in some way aware of his ϕ’ing x’ is not complex and does not seem hard to process. Since HO theorists can avail themselves of multiple forms of awareness, coming to terms with the fact that Carmichael’s ϕ’ing x seeming some way to Carmichael entails that Carmichael is aware of his ϕ’ing x does not seem to be much of a challenge.

The importance of this point should not be underestimated. Charity to understanding will force the HO theorist to re-interpret the FO theorist only if the HO theorist regards his own view as being discoverable without such difficult reasoning—only if, that is, he takes the truth of Transitivity to be not like other, more complex and difficult a priori truths. We have seen that given what the HO theorist means by ‘what it is like’-sentences, it is reasonable to conclude that she does see the truth of Transitivity that way. But we don’t actually need to make this inference on our own. For at least some HO theorists are forthright that they do see their view this way. William Lycan has been quoted (Gennaro 2011: 29) as wondering whether the HO approach—because of Transitivity itself—is “nearly trivially true.” Rosenthal describes Transitivity as “pre-theoretic” datum (1997a: 156). And Pessi Lyrra remarks that Transitivity is treated as a fundamentatum for further theorizing (2009: 68).

What these considerations show, I venture, is that as HO theorists, we should reject the assumption that FO theorists mean what we mean by Nagelian Transitivity and other ‘what it is like’-sentences and instead look for a more charitable interpretation. An obvious option is to

---

32 We will discuss Transitivity’s status as a pre-theoretic truism much further in § 3.
treat FO theorists as speaking FO-English. For then ‘what it is like’-sentences have their meaning in accordance with the Affective View, where a mental state is such that it is like something for one to be in it just so long as there is some way that one feels as a result of being in that mental state. The speaker of FO-English treats the ‘for’ in ‘what it is like for’ not as an operator that semantically expresses what ‘it seems to you that’ expresses, but as an argument of the verb phrase ‘is like,’ expressing affectiveness, i.e. some way the subject feels as a result of ϕ’ing. But now rejecting Nagelian Transitivity seems plausible. Suppose I feel an itch in my leg. It is true that I am thus aware (in a sense) of the itch. And it is true that the itch itself (i.e. the thing apparently in your leg) may itself be called a ‘feeling’ (Stoljar, forthcoming). So it is also true that if I have a feeling I am aware of that feeling. But that, as Stoljar rightly notes (ibid), doesn’t help the HO theorist. Indeed, it shows that HO theory is false; I am not aware of my awareness of the thing in my leg, despite the fact that I feel some way, and thus, by the Affective View, am in a conscious state.

Likewise on the ‘seems’ disjunct of the Affective View. Suppose Carmichael sees a Ming vase, and suppose that the Ming vase seems some way to Carmichael in virtue of his seeing the Ming vase. By the Affective View, Carmichael’s seeing the Ming vase will be conscious. But the Ming vase can seem some way to Carmichael without Carmichael being aware of his seeing the Ming vase, i.e. with Carmichael being aware of his awareness of the Ming vase.34

33 Jesse Prinz—a FO theorist—comes fairly close to this characterization (2012: 4).
34 Note that the Affective View does not rule out that, at least in some cases, I also represent the state as being a certain way (Stoljar, forthcoming). As understood here, FO theorists deny Transitivity, but doing so only implies that awareness of one’s mental state is not necessary for that mental state’s being conscious. It is consistent with that view that one can be (in some sense) aware of one’s conscious state. This is reflected in the Affective View; it is possible that I both feel a certain way in virtue of my ϕ’ing and represent my ϕ’ing as being a certain way.
2.3.4 Disputed and undisputed sentences, take one

Here is where we are. We began with a disputed sentence. This was Transitivity: ‘A mental state of which one is wholly unaware is not conscious.’ On FO theories Transitivity is false. On HO theories Transitivity is true. Yet both agree that Nagelian is true, so both agree that there are translations of the predicate ‘is conscious’ in terms of Nagelian. Thus Nagelian Transitivity is a disputed sentence as well. We also saw that the HO theorist ascribes to the Operator View. Finally, we argued that by employing a minimalist version of Hirsch’s framework, charity demands that treat the FO theorist as ascribing to something like the Affective View.35

The argumentation thus far has been conducted in fairly abstract terms. In effect, we argued that the FO theorist should agree that Transitivity is true in HO-English, and that the HO theorist should agree that Transitivity is false in FO-English. However, to make the case for Verbal truly compelling, whether they should is less useful than whether they would. Can we extract—from FO-English and HO-English, respectively—two actually undisputed sentences U1 and U2, one true and one false, such that the FO theorist holds that U1 is equivalent to Transitivity and the HO theorist holds that U2 is equivalent to Transitivity? I think we can.

35 Charity to understanding has done the brunt of our work. What about charity to retraction? The idea is that we should assume that speakers deal with new evidence in a reasonable way. As an interpretive norm this means favoring an interpretation that makes a communities’ rejections in the face of additional evidence come out right (Hirsch 2005: 74). So if someone were to recant a judgment about the behavior of falling bodies given new information—say, to use Hirsch’s example, the judgment that a ball thrown in the air at speed x will hit the ground at a speed much greater than x—we should, other things being equal, presume that the original assertion was false. While it’s unclear to me how much weight we should put on someone not retracting, I’ll simply note that I know of no rejections in the debate over Transitivity. In the exchanges recounted in § 2.2, when Rosenthal cites Nagelian as evidence that extinction subjects cannot be conscious, Block doesn’t retract. Indeed, Block doesn’t retract even though Rosenthal in effect conveys the belief that a failure to retract is crazy.

Verbal disputes in the theory of consciousness
My approach will have two parts. The first part sticks with our usual suspects, i.e. Block and Rosenthal. Technically, this will only entitle me to the claim that the dispute between Block and Rosenthal is verbal, rather than the claim that the dispute between FO and HO theorists is verbal, unless ‘Block v. Rosenthal’ is an illustrative case. Now various FO theorists differ from Block in various ways qua FO theorist and various HO theorists differ from Rosenthal in various ways qua HO theorist. But of course that does not matter. What matters is whether these are relevant differences—whether these differences will make it the case that $U_1$ and $U_2$ are disputed rather than undisputed. So, in case these differences are relevant, the second part (§ 2.3.5) looks at a way to recast the argument such that it generalizes to all FO and HO theorists.

To begin, let $U_1$ be ‘A mental state of which one is wholly unaware cannot be such that one feels some way in virtue of being in it.’ Let $U_2$ to be ‘A mental state of which one is wholly unaware cannot seem some way to one.’ Transitivity is equivalent to $U_1$ according to the FO theorist; Transitivity is equivalent to $U_2$ according to the HO theorist. I’ll now argue that Block and Rosenthal agree that $U_1$ is false and that $U_2$ is true.

The case for $U_1$ is trickier, so let’s look there first. We know that Block would agree that $U_1$ is false. Would Rosenthal? Recall again the earlier exchange between Block and Rosenthal. Rosenthal is discussing Block’s claim that states of visual extinction can be conscious. Rosenthal is incredulous. His incredulity is premised on Nagelian. Yet shortly thereafter we find him striking a more conciliatory attitude, distinguishing between two kinds of ‘phenomenality:’

One...consists in the subjective occurrence of mental qualities, while the other kind consists just in the occurrence of qualitative character without there also being anything it’s like for one to have that qualitative character. Let’s call the first kind thick phenomenality and the second thin phenomenality. Thick phenomenality is just thin phenomenality together with there being something it’s like for one to have that thin phenomenality (2002a: 657).
...Block’s view seems to be that phenomenality is simply thin phenomenality, and what I’m calling thick phenomenality is phenomenality plus reflexivity. For example, he seems to take the ability to report a mental state as an indication that reflexivity is present, presumably because reporting something indicates awareness of it (ibid).

What are we to make of this? Here are three things we know. First, ‘phenomenality’ is actually Block’s term: “What is phenomenality? What it is like to have an experience. When you enjoy the taste of wine, you are enjoying gustatory phenomenality” (2001: 202). Second, Rosenthal wants to demote the ‘phenomenality’ our mental states have in the absence of higher-order awareness to ‘thin phenomenality.’ Third, Rosenthal denies that states with mere thin phenomenality are like anything for their subjects. This shouldn’t come as a surprise. For given Rosenthal’s adherence to the Operator View, a mental state will be like something for its subject only if that mental state seems some way. And a mental state will seem some way to its subject only if that subject is aware of that mental state—precisely what is missing in states with thin phenomenality alone.

So what can we conclude? More specifically, what is it, by Rosenthal’s lights, for a mental state to have mere ‘thin,’ as opposed to ‘thick,’ phenomenality? Well, if ‘phenomenality’ is to have any purchase here whatsoever, it is plausible to think that what Rosenthal means is simply that while we can feel some way in virtue of being in states with thin phenomenality, those states are not truly conscious unless we are also aware of them. Yet then Rosenthal would certainly grant that U₁ is false: we can feel some way in virtue of being in a mental state without our being aware of that mental state. So U₁, it would seem, is indeed uncontested.36

36 One might think that Rosenthal’s use of the expression “qualitative character” settles things since he equates qualitative character with thin phenomenality, and qualitative character is often used interchangeably with phenomenal character. However, my guess is that this is not right; “qualitative character,” as Rosenthal uses the expression, figures in his Sellarsian quality space theory of perception (2005; 2010; 2015). On this view, there is an isomorphic mapping of mental qualities (or “qualitative
Here is some more evidence:

Bodily sensations such as pains can also occur without being conscious. For example, we often have a headache or other pain throughout an extended period even when distractions intermittently make us wholly unaware of the pain. (Rosenthal 2002b: 411)

One might say that this is evidence that Rosenthal would assent to $U_1$, only if we have reason to believe that, on Rosenthal’s view, one could feel some way in virtue of being in an unconscious pain. But unless Rosenthal has a highly idiosyncratic notion of sensations like pain, we can make the inference on Rosenthal’s behalf: pains are a type of feeling, so in having a pain one ipso facto feels some way. Given that “pains can...occur without being conscious,” Rosenthal would indeed likely claim that $U_1$ is false. Hence, once again, $U_1$ it would seem, is undisputed.

I see two objections that might arise here. To say that Rosenthal would claim that $U_1$ is false is not merely to say that there can be unconscious pains, but that there can be unconscious feelings—for Rosenthal, feelings that I am in no way aware of having. The first objection is that this is simply too implausible to take seriously. Whatever states Rosenthal intends to pick out by the term ‘pain’ must be such that those states are not feelings, i.e. states without a feel. Perhaps they are just subliminal first-order representations of tissue damage. (Naturally, the ‘too implausible to take seriously’ line, if pursued, would have to be pressed against Block, too; it is essential to FO theories that this is not just plausible, but actual).

character”) to sensible properties. The former represents the latter because this isomorphism holds (with the mental qualities extrapolated from the subject’s discriminatory capacities with respect to the sensible properties). Quality space theory is different from traditional representational. Here, perceptual contents—to borrow Berger’s phrase (2015)—are individuated holistically as opposed to atomistically, e.g. in terms of causal co-variation. Details aside, Rosenthal does grant that these qualitative characters have phenomenality, albeit thinly. So its hard to see—given how Block himself defines ‘phenomenality’—that we do not at least feel some way in virtue of being in such states.

VERBAL DISPUTES IN THE THEORY OF CONSCIOUSNESS
The second objection assumes that Rosenthal does recognize unconscious feelings, but points that this does not mean that Rosenthal would agree that the subject feels some way in virtue of being in such states. And this, the objection continues, is precisely what \( U_1 \) demands; by the Affective View, to say that \( U_1 \) is false, is to say that the subject can feel some way in virtue of being in a mental state, despite being unaware of that mental state. So we need to move beyond the merely sub-personal.

I’ll broach both worries in tandem. (The first worry will surface again in § 2.3.5). We might note right off the bat that whatever else we want to say about the implausibility of our having feelings, and thus feeling some way, despite not being aware of that feeling, it is not a mantle that Rosenthal would have to bear alone. For we already know that as a FO theorist, Block must recognize the possibility of such a case. But bracket this. Clearly there is a sense of the term ‘feeling’ where we—that is, subjects of experience—can feel some way and not be aware that we feel that way. Consider emotional feelings. I can be depressed or angry—I can feel depressed or feel angry—and not know that I am feeling these ways until a friend points out my aberrant behavior. Indeed, Rosenthal would concur; discussion of such cases is common in the literature on HO theories, meant to illustrate the requirement that the requisite higher-order awareness cannot come about by conscious inference (1997a; 2005). And critically, note that these attributions are person-level attributions; when your friend comes to you, she does not merely say that there is some sub-personal feeling, but rather that you feel some way. She does so because it is our feeling a certain way that explains our behavior. The HO theorist just says that we are not aware of the mental state in virtue of which we feel some way.
The upshot is that there is nothing about the notion of feeling as such that would make $U_1$ contested. This suggests an argument from parity: if we are already willing to make person-level attributions of emotions (a type of feeling) that we are unaware of, parity suggests that we should also make person-level attributions of pains (qua feelings) of which we are unaware.

We can press the point by fixing on a test case. Suppose I stub my toe. As a result I am in a state that represents my toe as having a throbbing quality. Call this $Q$. In the present context, to say $U_1$ is false is to say that I can instantiate $Q$ (or be in state with $Q$), and thus feel some way (the throbbing way), despite $Q$ (or my being in a state with $Q$) not seeming any way to me.\(^{37}\)

Would Rosenthal grant this possibility? In considering this, we need to remember that Rosenthal will not say that there is something it is like to (merely) be in a state with $Q$ because of what he thinks this phrase means. But that does not mean he would deny that we feel some way in virtue of being in a state that has $Q$. By comparison, consider the following:

It is beyond dispute that perceiving occurs not only consciously, but without being conscious as well, in masked priming and other forms of subliminal perceiving...Absent special assumptions about how to taxonomize mental qualities, the natural and indeed inevitable conclusion would be that, apart from the property of being conscious itself, whatever mental properties occur in conscious perceiving also occur in the subliminal, non-conscious case (2015: 36).

Although Rosenthal is concerned with perception here, what he says is telling. The idea is that apart from being conscious itself, whatever mental properties occur in conscious perceiving also occur in the non-conscious case. $Q$ is a mental property. Taking this point in conjunction with Rosenthal’s earlier concession that a state with $Q$ still has ‘thin’ phenomenality, a natural conclusion to draw is that we feel some way in virtue of being in a state with $Q$. After all, $Q$ is a

\[^{37}\text{Or more generally: despite my not being aware of } Q.\]
throbbing quality, i.e. a property that types states by how they feel. True, the state with $Q$ won’t seem any way, and so by Rosenthal’s lights won’t be conscious, until its targeted by a suitable state of higher-order awareness. Nonetheless, the state with $Q$ still has a throbbing quality, and it still has ‘thin’ phenomenality, and it has all these things absent being suitably targeted. But if a throbbing quality has *any* phenomenality, yet we deny that state with that quality seems any way, then its stands to reason that we must at least feel some way in virtue of being in that state. For otherwise, talk of being in state with throbbing quality and thin phenomenality loses all purchase. Thus, Rosenthal should agree that $U_1$ is false.\(^{38}\)

What if we switched to the disjunctive version of The Affective View, and focused on its second disjunct? Would that change things? Here, $U_1$ becomes $U_1^*$: ‘A mental state of which one is wholly unaware cannot be such that things seem some way in virtue of being in it.’ Would the HO theorist say that $U_1^*$ is false too? There is reason to think she would. Consider *semantic priming* in the lexical decision task (e.g. Dehaene et al 1998). Subjects are presented (usually visually, but auditorily as well) with a mix of words and pseudowords as targets preceded by non-conscious primes, which can either be semantically related (like *wolf* and *dog*) or unrelated (like *lawyer* and *turkey*) to the target words. That subjects are faster to respond to semantically related prime-target pairs relative to unrelated pairs suggests that a stimulus can be categorized and conceptualized without our being aware of seeing that stimulus. But if this is so, then it’s plausible that a stimulus can *seem* some way to us without our being aware of that we perceive that stimulus. Indeed, HO theorists (e.g. Weisberg 2008) are keen to agree. In general, when we

\(^{38}\) It also worth noting that for Rosenthal unconscious pains have the same causal profile (e.g. their role in avoidance behavior and wincing), as conscious pain (2012a). Although it might be conceptually possible, it is hard to see how unconscious pains could play this role without them feeling some way.
speak of seeing something as red (or as a turkey), and thus conceptualize that thing as red, that red thing must seem to us a certain a certain way—viz. the way red things typically seem. So, $U_1^*$, I venture, should be uncontested as well.

It is also worth noting, if only briefly, that the states under consideration here are wholly unlike those often posited in sub-personal psychological explanations. One hallmark of sub-personal states is that they often do not correspond to anything that we find in person-level explanations, even if the reverse is the case (Drayson 2014). For instance, take the internal grammatical states posited by Noam Chomsky to explain language learning in children, or the computational processes posited by David Marr in early visual processing. We cannot, as Zoe Drayson notes, “have beliefs about the complex mathematical equations that convert luminosity values into edges; and we can’t experience the contents of our stored grammatical rules, or use the information to draw inferences” (Drayson, ibid). Yet when HO theorists talk about day-long headaches or subliminal perception, these are states that we most certainly can have beliefs about or (in some sense) experience; indeed, they would have to be, for otherwise they could never be conscious by the HO theorist’s lights. Thus, it is hard to see how when HO theorists talk about feelings that we are unaware of, they are talking about sub-personal states.

What of $U_2$: ‘A mental state of which one is wholly unaware cannot seem some way to one’? We already know that Rosenthal would agree that $U_2$ is true. By the Operator View, a mental state can be like something for its subject unless that mental state seems some way to its subject. By Transitivity, a mental state cannot be like something for its subject unless its subject is in some

---

39 Remember: the FO theorist need only ascribe to one version of the Affective View. So if one does not think that things can seem some way to a subject without the subject being aware that things seem some way, one can always revert back to the non-disjunctive variant of the Affective View.
way aware of that mental state. So, that mental state cannot seem some way to its subject unless its subject is aware of that mental state. Hence, \( U_2 \).

Would Block assent to \( U_2 \)? I see no reason why he wouldn’t. It’s worth noting that what Rosenthal calls conscious mental states, Block calls conscious mental states *with reflexivity*—and he agrees that reflexivity is tantamount to higher-order awareness (2001).

Perhaps the FO theorist, on behalf of Block, might point out that whether \( U_2 \) is true depends on how we understand ‘seems’. It is well known that ‘seems’—like ‘appears’ and ‘looks’—have different uses (Chisholm 1957; Dretske 1969; Jackson 1977). Compare:

(1) **Non-Comparative**: The chair seems red.

(2) **Comparative**: The house seems old-fashioned.

(3) **Epistemic**: It seems to be the case that Carmichael was sleep walking again.

On the ‘non-comparative’ use of ‘seems’ in (1), ‘seems’ is functioning as a perceptual verb. It seems clear enough that if the chair seems red to a subject, then that subject must be aware of the chair as being red. Likewise on the ‘comparative’ use of ‘seems’ in (2): for the house to seem like an old-fashioned house, I must be aware of the house.

The epistemic ‘seems’-statement in (3) is trickier. When ‘seems’ is used epistemically, the sentence conveys what is subjectively probable conditional on some body of evidence (Brogaard 2014), amongst others, has claimed that all semantically non-comparative (or ‘phenomenal’) ‘seems’ statements (and other ‘looks’-statements) are perceptual. But I doubt this; as Brogaard herself admits, ‘seems’ and ‘looks’-statements can be individuated in terms of the mental state they claim to be based on. So, while we can have perceptual ‘seems’-statements like (1), we can also have *introspective* ‘seems’ statements, such as ‘My migraine seems bad.’ Yet introspective ‘seems’ statements will only be undergirded by a process that is perceptual in nature if one adopts an inner-sense account of introspection. In addition, I don’t see any reason why a HO theorist could not say that the relevant sense of ‘seems’ is phenomenal; the appearances conveyed by the state of higher-order awareness (e.g. what that state represents) can phenomenologically seem some way without that state itself being an object of higher-order awareness (and hence, by the HO theorist’s lights, conscious itself).
So Anthony might utter the sentence in (3) because upon walking into the kitchen in the morning, he finds food all over the floor, and his wife’s purse in the refrigerator. In such a case, Anthony is certainly aware of the food on the floor and his wife’s purse. But note that although this may cause Anthony to be aware that Carmichael was sleepwalking, one might deny all the same that Anthony is aware of Carmichael; one might, as we’ll see in § 3.4.4, deny that de dicto awareness entails (direct) de re awareness (e.g. Dretske 1995; Tye 2014).

This complication aside, I highly doubt that any HO theorist will have the epistemic sense of ‘seems’ in mind. What it is like for me to have a conscious experience does not change when faced with evidence contrary to the deliverances of that experience. If I have an experience where it seems to me as if there is a tiny rainbow colored elephant in my visual field, it will continue to seem this way even if I come to find out, on independent grounds, that I had been hallucinating all along. Epistemic seemings, by contrast, will quickly ebb in presence of defeater, at least if the subject is rational (Brogaard 2014).

So $U_2$ should be undisputed as well. But if $U_1$ and $U_2$ are both undisputed, it follows that the dispute over TRANSITIVITY is verbal. In FO-English, TRANSITIVITY is false because it is equivalent to the (undisputedly false) sentence $U_1$; in HO-English, TRANSITIVITY is true because it is equivalent to the (undisputedly true) sentence $U_2$. The malefactor is the phrase for the subject. FO and HO theorists disagree about TRANSITIVITY, at bottom, because they disagree about how to interpret ‘for the subject’ in ‘what it is like’-sentences.

2.3.5 Disputed and Undisputed Sentences, Take Two

I see two interconnected threats to this argument, both of which focus on HO theory and HO-English. Our initial set up began with a disputed sentence (D₁), which just was Transitivity:

\[ D₁: \text{ A mental state of which one is wholly unaware is unconscious.} \]

HO theorists think that D₁ is true; FO theorists think that D₁ is false. We also had an undisputedly true sentence (UNAG), which was just Nagelian:

\[ UNAG: \text{ For any mental state } M \text{ of a subject } S, M \text{ is conscious iff } M \text{ is such that there is something it is like for } S \text{ to be in } M. \]

Finally, D₁ and UNAG together yielded another disputed sentence (D₂), Nagelian Transitivity:

\[ D₂: \text{ A mental state of which one is wholly unaware cannot be like something for its subject.} \]

The argument that the dispute over D₁ and D₂ is verbal turned on a further pair of undisputed sentences, stemming from two languages: HO-English (where ‘what it is like’-sentences have their meaning in accordance with the Operator View), and FO-English (where ‘what it is like’-sentences have their meaning in accordance with the Affective View). They were:

\[ U₁: \text{ A mental state of which one is wholly unaware cannot be such that one feels some way in virtue of being in it.} \]

\[ U₂: \text{ A mental state of which one is wholly unaware cannot seem some way to one.} \]

The claim, then, was that while U₁ is undisputedly false and U₂ is undisputedly true, the FO theorist claims that U₁ is equivalent to D₁, whereas the HO theorist claims that it is actually U₂ that is equivalent to D₁. This renders the dispute between HO and FO theories verbal.

However—and here’s the worry—the case for the claim that U₂ is undisputedly false rested on an interpretation of Rosenthal’s HOT theory. This raises two potential problems. The first is
that this interpretation is just wrong: despite talk of unconscious headaches and mental states having ‘thin’ phenomenality despite not being HOT-targeted, Rosenthal would just deny that we will feel some way in virtue of being in a mental state unless that mental states seems some way—i.e. unless one is aware of that mental state. The second problem brackets the question of what Rosenthal says and focus on HO theory itself: is it the case that HO theory as such must say that we can feel some way in virtue of being in a mental state despite that mental state’s not seeming any way to its subject? If the answer is no, then our argument fails to generalize. And maybe that’s what a HO theorist should say, even if that’s not what Rosenthal says.

In response, I think there is way of reworking our initial setup in § 2.3.4 that can preserve the verdict that the dispute between HO and FO theorists is verbal even if the aforementioned reading of Rosenthal’s approach is wrong, and even if HO theory as such is not committed to the falsity of U₁. To see this, suppose we introduce two new undisputedly true sentences:

\[ U₃: \text{For any mental state } M \text{ of a subject } S, \text{ if } S \text{ feels in } M, \text{ then } M \text{ is conscious.} \]

\[ U₄: \text{For any mental state } M \text{ of a subject } S, \text{ if } S \text{ feels in } M, \text{ then } M \text{ is like something for } S. \]

Now Rosenthal, as claimed in § 2, would say that U₃ and U₄ are actually false. But this version of Rosenthal—call him ‘Rosenthal*’—would not. Rosenthal* thinks that feeling some way is sufficient for being in a conscious state. Thus when Rosenthal* talks about Carmichael being unaware of his headache and thus his headache being unconscious (by TRANSITIVITY), Rosenthal* will also deny that Carmichael feels anyway in virtue of having the headache. For if he did not, Rosenthal* would then be forced to say that U₃ is false. Moreover, Rosenthal*, like almost all HO and FO theorists, still ascribes to NAGELIAN. So if a subject’s feeling some way is sufficient for a
subject’s being in a conscious state \( (U_3) \), then a subject’s feeling some way is sufficient for there being something it is like for the subject—hence \( U_4 \).

However, Rosenthal and FO theorists like Block would still disagree over the following sentence:

\[ D_3: \text{ For any mental state } M \text{ of a subject } S, \text{ if } S \text{ feels some way in virtue of being in } M, \text{ then } M \text{ seems some way to } S. \]

Rosenthal claims that \( D_3 \) is true. Indeed, he must: Rosenthal is like Rosenthal (and seemingly all HO theorists) in that they both ascribe to the Operator View. On that account, the meaning of ‘\( M \) is like something for \( S \)’ in \( U_4 \) translates to ‘\( M \) seems some way to \( S \)’ in \( D_3 \).

By contrast, FO theorists like Block say that \( D_3 \) is false because they reject TRANSITIVITY: a subject \( S \) can feel some way in virtue of being in \( M \) despite the fact that \( S \) is not aware of \( M \). And if \( S \) is not aware of \( M \), \( M \) cannot seem anyway to \( S \). Yet this does nothing to undermine the claim that FO theorists reject TRANSITIVITY because of a fundamental disagreement with HO theorists over the meaning of ‘what it is like’-sentences. For if they ascribe to the Affective View—and by the arguments given in § 2.3.3, it is only charitable to assume they do—then the ‘for \( S \)’ in \( U_4 \) will only serve to make explicit the subject who is affected in a particular way in virtue of being in \( M \)—a state of seeing a Ming vase, say. It won’t, in other words, function as an operator, expressing how \( M \) itself seems to the subject.

2.4 **Prolegomenon to a Future Science of Consciousness**

So the dispute between all FO and HO theorists—not just the dispute between Block and Rosenthal—is verbal. But it is, all the same, substantive. To see this, recall the two models for assessing the metaphysical merit of competing languages mentioned earlier (§ 2.3.1). The first
model—advocated by Sider—said that metaphysical merit of two languages \( L \) and \( L^* \) is a function of (i) the joint-carving status of the terms of \( L \) and \( L^* \) and (ii) relative the intensional resources of \( L \) and \( L^* \). The second model—advocated by Hirsch—focused on (ii) alone. The predicate ‘green’ captures similarity in a way that ‘grue’ doesn’t.\(^{42}\) Green things are more similar than grue things and thus they mark a more natural joint in nature. So a language that employs ‘green’ as opposed to ‘grue’ will not be metaphysically on par precisely because the former but not the latter captures something more fundamental about reality. It is true that ‘\( M \) seems some way to \( S \)’ (HO-English) and ‘\( S \) feels some way in virtue of being in \( M \)’ (FO-English) are, prima facie, nothing like ‘green’ and ‘grue’. But for reasons adduced earlier (§ 2.3.1), the Siderian epistemology of structure makes deciding this issue beyond the scope of our project.\(^{43}\)

So we are left with intensional resources. Compare our present case with disputes in physical object ontology. Suppose we have an anti-nihilist and a nihilist. The anti-nihilist and the nihilist speak different ontological languages, individuated by which quantificational sentences are true in that language (Hirsch 2011, xiv). So if \( \exists x Fx \) is true in \( L \), and \( \neg \exists x Fx \) is true in \( L^* \), then \( L \) and \( L^* \) are different ontological languages. But despite being different ontological languages—despite being different with respect to what they say there is—\( L \) and \( L^* \) are truth-functionally equivalent; ‘There are tables’ (from \( L \)) and ‘There are simples arranged tablewise’ (from \( L^* \)) are true with respect to the same possible worlds. To use John Hawthorne’s expression (2009: 220), the anti-nihilist makes no intensional advance in speaking \( L \) over the nihilist who speaks \( L^* \). There

\(^{42}\) The notion of naturalness and structure as applied to predicates stems from Lewis (1984; 1986).

\(^{43}\) For all we know, it might be the case that the category of things that can feel some way in virtue of being in a mental state is the same as the category of things whose mental states can seem some way, even if these properties are distinct. That strikes me as an empirical question.
is no true intension that she accepts that the nihilist denies. The languages of the nihilist and anti-
nihilist thus have the same fact-stating, expressive power.

When we look at FO-English and HO-English, however, things are different. It is not true
that, for any sentence \( s \) in FO-English, a speaker of HO-English can, in the same context, find a
sentence \( s^* \) in HO-English which is true with respect to the same possible worlds. How so? Well,
when the HO theorist denies that Carmichael is in a conscious state, what she is denying is that
that mental state seems some way to Carmichael. Thus what is needed is some sentence in FO-
English that, relative to any (actual or possible) context of utterance, is true in the same worlds
as ‘Carmichael’s mental state seemed some way to Carmichael’. But the sentence offered by the
FO theorist—‘Carmichael felt some way in virtue of being in a mental state’—is clearly not true
with respect to the same worlds. There will be worlds where Carmichael felt some way in virtue
of being in a mental state, but where that mental state didn’t seem any way. Or, if one wants to
employ the disjunctive version of the Affective View, there will be worlds where things seemed
some way to Carmichael in virtue of being in a mental state, without that mental state itself
seeming any way. So, FO and HO-English are not intensionally equivalent. Hence, FO-English and
HO-English are not of equal metaphysical merit.

It is worth noting that there is yet another way in which the dispute over TRANSITIVITY
remains substantive, viz. that there are further questions that clearly are substantive, which
plausibly turn on whether TRANSITIVITY is true. For instance, there are tough moral questions
surrounding how we should assess the value of the lives of subjects who are in persistent
vegetative states. Whether they have conscious states might be relevant to this assessment. If it
is lack of awareness that makes their mental states unreportable, then whether conscious states
are states we are conscious of is likely relevant to how we assess the value of the lives of subjects who are in persistent vegetative states. 44

Therefore, the dispute over TRANSITIVITY is both verbal and substantive. 45 I know of no one in the literature who has seriously entertained the thesis defended here. (I say ‘serious’ because, as noted in the epigraph that opens § 2, Rosenthal himself flirts with the idea, but presumably in jest). Kriegel (2009: 3-5) comes the closest. Kriegel argues that although the dispute between FO and HO theorists might seem verbal, it is ultimately not, because each side disagrees about which property is responsible, in the actual world, for the mystery of consciousness—something that can, and ought, to be gotten right. I think Kriegel is conflating verbalness with non-substantiveness, so my story might actually be compatible with his. I just think that each side is targeting different properties because they mean different things by ‘is conscious’, and that the dispute is verbal because if they meant the same thing they would actually agree. If Kriegel just means verbal—i.e. verbal without the further claim of non-substantiveness—then my argument undercuts his claim.

But now we are left with a critical question: how do we proceed, if the dispute over the dispute over TRANSITIVITY is substantively verbal? This issue is critical since whether conscious

44 See Chalmers (2011) and especially Greco (2015) for a similar construal of substantiveness. The basic idea here is what Greco calls a transmission model of substantiveness, according to which the substantiveness of one question will hinge on its relation to other questions that are substantive. (So if you think that the question of the moral status of persons is a substantive question, and you think that whether something is or can be conscious bears on whether something is a person, then disputes over when something is conscious is itself a substantive matter).

45 Of course, this does not imply that the individual questions FO and HO theorists are asking are themselves non-trivial. For example, in HO-English the question HO theorists are answering when asked whether TRANSITIVITY is true is whether a mental state can seem some way without our being aware of that mental state. There might be a sense in which this claim is trivially, but that is neither here nor there; contra Jackson (2014: 45), that a dispute is verbal, does not mean that neither side’s claims are trivial or non-substantive (cf. Hirsch 2013: 437).
states are states we are aware of, and by extension, whether a FO or HO theory is true, are amongst the central questions in the philosophical and scientific study of consciousness.

In general, when parties come to realize that their dispute is verbal, they can do one of two things. The first option is to stop talking about the topic of that dispute. That seems like the appropriate reaction when the dispute is non-substantively verbal. When participants to such a dispute realize that it is verbal, they simply direct their attention elsewhere, or at the very least that’s what they ought to do. It is simply not worth our time. The dispute I have over questions such as ‘Is a glass a cup?’ fall into such a category because how we answer them doesn’t matter; they turn on which of a range of equally fine available meanings (e.g. ‘receptacle made for drinking’) we choose for the expressions (e.g. ‘glass’) in those questions (Sider 2012: 54).

The second option is to continue talking about the topic of the dispute, but to change the way we approach the topic of the dispute, or the terms in which that dispute is conducted. This is the appropriate reaction when the dispute is substantively verbal. In a dispute that is verbal but substantive, the dispute is fundamentally about language, but that dispute itself matters. So to adjudicate the dispute we need to settle questions about language.

How does this play out in the present case? To settle the dispute over TRANSITIVITY, it is common to appeal to the results of the cognitive and neurological sciences—for instance, by Sperling paradigm cases (Block 2007; Brown 2012a), or lesion evidence in the pre-frontal cortex (Lau & Rosenthal 2011; Kozuch 2014). But if the dispute is substantively verbal, these issues are best placed downstream; what we need to do, first and foremost, is appeal to the results of the

---

46 On the latter, the thinking is that if some form of HO theory is true, then lesions to the pre-frontal cortex should result in deficits in consciousness, because the pre-frontal cortex is widely thought to be the source of states of higher-order awareness.

VERBAL DISPUTES IN THE THEORY OF CONSCIOUSNESS
science of *language*. Four points help bring this out. First, FO and HO theorists alike agree that NAGELIAN is true. Second, the dispute over TRANSITIVITY is verbal because they disagree about the meaning of the expression ‘for the subject’ in ‘what it is like’-sentences like NAGELIAN TRANSITIVITY. Third, one can have a verbal dispute in which one side is misusing a portion of language that the other side is using correctly. Finally, because the dispute remains substantive, there is a privileged description of ‘is conscious’ in terms of NAGELIAN. So to figure out whether TRANSITIVITY is true or false, we need get a better grip on the semantics of ‘what it is like’-sentences.\(^{47}\) Of course, it might be the case that—contra our current assumption—NAGELIAN fails to track *anything* interesting about the nature of consciousness. But barring that, what emerges is a sort of prolegomenon to Stoljar’s research. I haven’t endorsed the Affective View—and nor have we discussed Stoljar’s case for it—but whether we should is precisely the sort of question we need to ask if the considerations presented above are correct.

---

\(^{47}\) I don’t mean to imply that this is the only appropriate response. Perhaps there might be other available shifts in research program that don’t involve only appealing to linguistics. For instance, there may be non-linguistic questions about which use of ‘consciousness’ is best given our general practices, and our desire to explain certain behaviors. But notice that even on this option the way we have been going about trying to decide on TRANSITIVITY is (at least partially) wrong. While I do not intend to argue for my favored option here, it seems clear enough that, given our current thesis, taking a hard look at the semantics of ‘what it is like’-sentences is inevitable if we hope to make headway on TRANSITIVITY.


3 TRANSITIVITY AND TRANSPARENCY

3.1 TWO PURPORTED TRUISMS ABOUT CONSCIOUSNESS

The central topic of the last chapter was the following thesis:

**TRANSITIVITY:** Conscious mental states are mental states we are aware of in some way.48

There we asked a second-order question: why do HO and FO camps disagree about TRANSITIVITY? Here we move to a first-order question: what is the status of TRANSITIVITY itself?

TRANSITIVITY occupies a curious place in the philosophical literature on consciousness. On the one hand, it is, as mentioned, a dividing point between two major theoretical families: HO theorists, who accept it, and FO theorists, who reject it. On the other, HO theorists believe that TRANSITIVITY is *obviously* true. We noted this as well in passing earlier (§ 2.3.3). Lycan has been quoted as wondering whether TRANSITIVITY is “nearly trivially true” (Gennaro 2011: 29). Rosenthal describes TRANSITIVITY as a “pre-theoretic” datum (1997a: 156) and as “intuitively obvious” (2000: 273). And Lyrra describes TRANSITIVITY as a *fundamentatum* (2009: 68).

I think the best way to summarize these sentiments is like this: for the HO theorist, TRANSITIVITY is a *truism*. A truism is a proposition that is not only deemed true, but also deemed obviously true. Truisms carry methodological import. Truisms can be assumed at the beginning of inquiry. We theorize *from* truisms, not to them. For the HO theorist, the upshot is significant. For HO theories are simply implementations of TRANSITIVITY. So to say that TRANSITIVITY is a truism

48 Here we switch back from the contrapositive of the thesis. Also, it is worth noting explicitly now that TRANSITIVITY is sometimes framed as the claim that conscious states are states we are conscious of (e.g., Rosenthal 2005: 3). I’ll continue to assume that ‘conscious’ and ‘aware’ are synonymous, and use them interchangeably. Note that this does not make expressions like ‘conscious awareness’ redundant; one can be aware of something without being conscious that one is aware.
is just to say that HO theory is a truism, i.e. that (some form of) HO theory is obviously true. So, for the HO theorist, the debate with FO theorists must seem frivolous.

Here’s another allegedly intuitive idea:

**TRANSPARENCY:** Our conscious states, or conscious experiences, are in some way *transparent* to awareness.\(^{49}\)

TRANSPARENCY, or something near enough like it, enjoys widespread support. Most FO theorists and many HO theorists adopt it.\(^{50}\) I am hesitant, however, to say that TRANSPARENCY is a truism, at least as it stands. I have two reasons. First, while many enthusiastically endorse the claim that experience is transparent, it’s an open question whether the enthused are actually enthused about the same phenomenon (Schroer 2007). Second, even assuming the enthused have a bead on a single genuine phenomenon, the scope of this phenomenon is another source of widespread debate. Very few reject TRANSPARENCY outright; in most cases, it is the *universality* of TRANSPARENCY that is rejected.

Even if TRANSPARENCY is not itself a truism, we can readily extract a truism from TRANSPARENCY. Doing so consists of two steps. The first is a disjunctive move. The basic idea behind TRANSPARENCY is that there is an awareness-relation that we seemingly cannot bear to our experiences. But the nature of this awareness-relation is a matter of controversy. By treating this awareness-relation

\(^{49}\) It is perhaps more common to talk of the transparency of *experience*, than the transparency of conscious states. But this substitution is harmless (cf. Wright 2005). As I use the term, we can (and I will) move easily between ‘conscious state’ and ‘conscious experience’, since an experience is an event that consists of the instantiation of a conscious state. We can think of conscious states as properties, and experiences as events akin to the kind described by, say, Jaegwon Kim’s (1991) theory of events (cf. Bourget [2010] for this usage). HO theorists also often use ‘experience’ to pick out unconscious mental states (e.g. Carruthers 2000; Rosenthal 2005), so ‘conscious experience’ is not redundant on their usage.

\(^{50}\) An incomplete but representative list of sympathizers includes: Moore 1922; Harman 1990; Dretske 1995; Martin 2002; Tye 2002; Speaks 2009; Schroer 2007; Gennaro 2003, 2008; Kriegel 2009.
as a disjunction of awareness-relations, however, we can mitigate that controversy. The second is a restrictive move. By restricting the class of things that are transparent to largely uncontroversial cases, we mitigate the controversy surrounding the scope of transparency. Call the resultant thesis ‘weak’ transparency, or W-TRANSAPRENCY for short. W-TRANSAPRENCY is, if not a truism in its own right, near enough for our sake.

I have two aims in this chapter, one primary and one secondary. My primary aim is to investigate a puzzle facing the HO theorist. On the one hand, we have the purportedly pre-theoretic truism, advanced by all HO theorists, that all conscious states are states we are aware of in some way. On the other hand, we have another purportedly pre-theoretic truism according to which we are in some way unaware of at least some conscious states. TRANSITIVITY and W-TRANSAPRENCY both include ‘in some way’ clauses. I am interested in whether there is a precisification of TRANSITIVITY’s ‘in some way’ clause that is (i) compatible with the various ways of precisifying W-TRANSAPRENCY’s ‘in some way’ clause, yet (ii) continues to motivate HO theory. I’ll argue that there is no such precisification.

This result is important. For one thing, even if FO theorists reject it, TRANSITIVITY is all the same undeniably extremely intuitive; it is certainly odd to speak of a conscious state of which we are wholly unaware. So it would be nice, given their seeming pre-theoretic intuitiveness, if we could consistently ascribe to both TRANSITIVITY and W-TRANSAPRENCY. If I am right, we cannot, or at least not if our precisification of TRANSITIVITY is robust enough to sustain a HO theory. Second, and relatedly, several HO theorists (e.g. Gennaro 2003, 2008; Kriegel 2009) have explicitly argued that

---

51 This is to be distinguished from what Kind (2003) calls “weak transparency”.

TRANSITIVITY AND TRANSPARENCY
TRANSITIVITY and TRANSPARENCY (and by extension, W-TRANSPARENCY) are consistent. So if I am right, these HO theorists are wrong.

My secondary aim concerns whether TRANSITIVITY and W-TRANSPARENCY really ought to be treated as truisms. I’ll suggest (§ 3.5) that if TRANSITIVITY and W-TRANSPARENCY do conflict, then they ought not to be treated as truisms. This conclusion, however, does not require that TRANSITIVITY and W-TRANSPARENCY be ultima facie incompatible. All it requires is something weaker—that it is extremely difficult to see how TRANSITIVITY and W-TRANSPARENCY could be compatible. If it is extremely difficult to see how one purportedly obvious thesis could be compatible with another purportedly obvious thesis, then it is imprudent to act as if either thesis can be treated as a fundamentatum in our theorizing about consciousness. Given that both of these theses are often treated in this way, this result is important as well, for it implies that there is something wrong with the way we go about theorizing about consciousness.

3.2 GETTING TRANSPARENT ON TRANSPARENCY

The literature on transparency is large. It is also a mess. Thus, some regimentation will be useful. Let a transparency thesis be a description of the transparency phenomenon. David Chalmers (2013: 350) points out that a transparency thesis can be schematized like so:

**SCHEMA** [for a transparency thesis]: “It is (i) difficult/impossible to stand in (ii) a certain relation to (iii) certain mental entities.”

SCHEMA has three dimensions of variation: (i), (ii), and (iii). The dimension of variation involved in (i) is between what we’ll call prima facie and ultima facie transparency (Kind 2003). The relation in (ii) may be attention, direct attention, awareness, direct awareness, introspection, and so on. Dimension (iii) has been said to include intrinsic features of experience, the phenomenal
character of experience, and conscious states (experiences) themselves. For now, the point to note is this: there has been a transparency thesis for each permutation of SCHEMA. As I said: a mess.

To clean up, I propose two moves. We’ll do the restrictive move first. This move actually has two components. The first is a restriction of cases. No matter how we fill in SCHEMA, an unrestricted transparency thesis will be controversial [see, e.g. phosphine experiences (Block 1996); blurred vision (Smith 2008); bodily sensations (Kind 2003)]. Fortunately, we do not need an unrestricted transparency thesis, since TRANSITIVITY is unrestricted in scope. However we fill in (iii)—whatever it is that we cannot (or have difficulty) stand(ing) in a certain relation to—it is sufficient for our purposes to restrict the scope of (iii) to the least controversial cases. So this means bracketing tricky cases like bodily sensations and atypical visual experiences.52

The second restriction consists of treating transparency as a purely negative phenomenon. Discussions of transparency often invoke the metaphor of a pane of glass (see, e.g. Kind 2003: 226). Characterizing transparency as a purely negative phenomenon amounts to characterizing the nature of the glass (i.e. what role the glass is supposed to play in the metaphor), and characterizing the relation we evidently don’t bear to the glass. Characterizing transparency as a positive phenomenon amounts to characterizing what is revealed to us beyond the glass. For instance, Harman’s famous case of Eloise is not merely a negative illustration, e.g. a case of not being aware of our experiences. It is also a positive illustration; the colors Eloise experiences are experienced as features of the “tree and its surroundings” (1990: 39). Similarly, Tye says that, at

52 In fact, some HO theorists support transparency in unrestricted form (Gennaro 2008: 50; Kriegel 2009: 182). So we are assuming something weaker than what HO theorists already take on.
least in veridical cases, all we are aware of are objects (and properties of those objects) in our ambient environment (2014: 40; cf. Speaks 2014). These claims are based on a reasonable inference from the way things seem. If it seems like what we are aware of are just features of the “tree and its surroundings,” then—absent extremely compelling reasons to think otherwise—that is what we are aware of. Still, by dropping the positive claim, our understanding of transparency becomes less restrictive. When I have a visual experience of a red apple, the redness that constitutes what it is like for me may be instantiated by an amalgam of sense data. Or, it may just be a represented property of a physical apple. It is unlikely that a HO theorist will be happy with the former, but it is nice to know that there is nothing presumed here that precludes her doing so.

The second move concerns dimension (ii) of SCHEMA, but let’s detour briefly to (i) and (iii). On (i): we want ultima facie transparency, in line with the vast majority of transparency theses. Prima facie transparency obviously does not conflict with TRANSITIVITY and so is not of issue here.

On (iii): we want transparency of conscious states (or experiences).53 To be clear, transparency of conscious states does not mean that the phenomenal character of our conscious

53 Another option found in the literature—perhaps due originally to Harman—is transparency of intrinsic qualities to [fill in the relation found in (ii) of SCHEMA]. I’m going to bracket this option, so let me explain why. First, Harman notwithstanding, the point about intrinsicality is now often dispensed with as extraneous in discussion of transparency (e.g. Stoljar 2004; Nida-Rümelin 2007; Tye 2014). Second, the notion of experiencing a feature as intrinsic strikes me as obscure. (Even when we consider what we experience weight as—which some say seems intrinsic despite not being intrinsic—I’m not sure what this means). But putting these two issues aside, I actually don’t think the focus on intrinsic features matters. Whatever ‘features’ we are aware of in an experience, be they extrinsic or intrinsic, the point is that they seem to be ‘features’ (if anything) of objects located in our ambient environment. If being aware of a feature of x is a sufficient condition for being aware of x, then the fact that all the features we are aware of seem to be features of things other then our conscious states is good evidence that we are (in the relevant sense of ‘aware’) not aware of our conscious states.
states is transparent to awareness. This construal is sometimes floated, but as Speaks (2014) and Tye (2009, 2014) make clear, this is a misreading, and an implausible one at that.

In addition, this construal has a strong precedent, and like ultima facie transparency, it has the benefit of putting the tension with Transitivity into sharp relief.\footnote{For formulations that take this route, see Wright (2005), Gennaro (2008) and Tye (2002, 2009, 2014).} Transitivity is in the first instance a claim about conscious states. By this the HO theorist does not merely mean that we are aware of what those conscious states represent. Indeed, she cannot mean this. Per HO theory, the way we are conscious of our conscious states is fixed by the content of the state of higher-order awareness. States of higher-order awareness have higher-order content. Higher-order content is content that involves representations or representational properties; even if conscious states are themselves represented, they are still representations, and hence (also) what does the representing. If we were to parse Transitivity as the claim that we are merely aware of what our conscious states represents—things like Eloise’s tree—states of higher-order awareness could not have higher-order content, a manifest absurdity. Higher-order states might also represent what our (first-order) conscious states represent (e.g. redness, trees), but it cannot only represent these things. So while we can describe Transitivity as the claim that conscious mental states are mental states we are aware of as (say) having certain contents, this cannot be shorthand for the claim that we are just aware of what contents our conscious states have. A FO theorist like Tye or Dretske could accept that reading of Transitivity, as there is nothing here that requires a state of higher-order awareness.\footnote{Relatedly, if we took that reading, there would then be no straightforward argument from Transitivity to HO theory, contra what its proponents suggest (see, e.g. Lycan 2001; Kriegel 2009). See also § 3.5.} So, there is no HO theory-compatible implementation of Transitivity on which we are merely aware of the contents of our (first-order)
conscious states. We are also aware (in some sense) of the conscious states themselves, and their representational properties (e.g. the property of representing red). I discuss ‘HO theory-compatible’ readings in § 3.3.

So much for (i) and (iii). Finally, regarding dimension (ii), the relation cannot be awareness tout court (Stoljar 2004). That transparency thesis is implausible. It also renders any question of conflict with TRANSITIVITY a non-starter. The relation will thus be a specific type of awareness. The issue is what type. This is where the second, disjunctive move emerges: we simply fill in (ii) with a disjunction. Thus, collating our earlier points, we have:

**W-TRANSPARENCY:** For at least one conscious state M, it is impossible to:

(a) **TRANSPARENCY**

(b) **TRANSPARENCY**

(c) **TRANSPARENCY**

where ‘W,’ again, stands for ‘weak.’ What makes it weak is that it is defined disjunctively, negatively, and confined to relatively uncontroversial cases. I submit that this transparency thesis is a truism. It strikes me as at the very least no less plausible than TRANSITIVITY, and so a conflict with TRANSITIVITY should give the HO theorist pause.

Critically, **W-TRANSPARENCY** leaves the HO theorist daylight. Take **TRANSPARENCY**. That we can’t stand in a direct awareness relation to a conscious state does not preclude standing in an indirect awareness relation to a conscious state. To comport with **TRANSPARENCY**, then, it must be possible to precisify TRANSITIVITY in terms of indirect awareness. Or **TRANSPARENCY**: that we can’t stand in a de re awareness relation to a conscious state does not preclude being aware that

---

56 For examples in the literature of commitments to (a) see Tye (2002) and Stoljar (2004), for (b) see Tye (2009, 2014), and for (c) see Kind (2003) and Kriegel (2009), amongst others.
we are in a conscious state. So to comport with Transparency$_{de\ re}$, it must be possible to precisify Transitivity in terms of de dicto awareness.$^{57}$

What about Transparency$_{int}$? Transitivity says nothing about introspection on its own terms; it just says that we are aware of our conscious states. And in principle one could affirm Transitivity while denying that there is such a thing as introspection. Of course, such a move should only come as a last resort. So the real question is whether it is possible to precisify Transitivity in a way that allows for us to be aware of our conscious states, even though they are not introspectable. If that is possible, than Transitivity is compatible with Transparency$_{int}$. To most, this route might sound like a non-starter; if one is aware of $x$, $x$ must be introspectable, at least in principle. But since, as we will see, one HO theorist (Uriah Kriegel) denies precisely this, it is worth exploring.

Before we press on, a quick objection to dispense with: why not include in our disjunctive treatment of (ii) in Schema an attention relation, i.e. why not add:

(d) Transparency$_{attend}$: Stand in a relation of attention to $M$

to (a) – (c)? After all, its not as if Transparency$_{attend}$ is uncommon. And moreover, awareness might outstrip attention; even if I cannot attend to something, this doesn't mean that I cannot be aware of it. So if only Transparency$_{attend}$, is true, Transitivity is in the clear.

Two points are worth making here. First, the claim that (de re) awareness outstrips attention is controversial. For example, Tye argues that if one can form a de re cognitive attitude about something—as would be required on, say, any HOT-theoretic precisification of

$^{57}$It is possible that some awareness reports—like ‘$S$ is aware o is $F$’—are ambiguous between de dicto and de re readings. On the standard view, whether the report ‘$S$ is aware o is $F$’ picks out de re awareness depends on whether it permits substitution salva veritate of a co-referring term for ‘o’, whether it permits exportation and quantifying in. So if ‘$S$ is aware o is $F$’ is true, then awareness is de re only if ‘$S$ is aware o* is $F$’ is true, where o* = o, and only if ‘$\exists x$ (S believes that x is $F$)’ is true.
Thus if a HO theorist were to accept \textsc{transparency}_\text{attend}, she may be forced to give up \textsc{transitivity} altogether. Second, even if Tye is wrong, we can consider \textsc{transparency}_\text{attend} at least indirectly. Since \textsc{transparency}_\text{int} does not come with any restrictions about what introspection \textit{is}, the HO theorist is open to treat introspection as a form of attention itself. And we’ll look at both types of cases: ones where the HO theorist does not, or does not necessarily, assimilate introspection to attention (§ 3.4.1), and one where she does (§ 3.4.2). So \textsc{transparency}_\text{attend} will have its day after all.

\section*{3.3 Acceptable Consistency Glosses}

Let a \textit{consistency gloss} on \textsc{transitivity} be a precisification of \textsc{transitivity} that renders \textsc{transitivity} consistent with \textsc{w-transparency}. A precisification of \textsc{transitivity} is a \textit{consistency gloss}, then, if it is consistent with \textsc{transparency}_{\text{direct}}, \textsc{transparency}_{\text{de-ref}}, or \textsc{transparency}_{\text{int}}.

Consistency comes in two forms. The first is unmitigated consistency: a precisification of \textsc{transitivity} is consistent, in an unmitigated sense, with \textsc{transparency} if, on it, we are \textit{not} aware of our conscious states (in the relevant sense). The other is phenomenological: a precisification of \textsc{transitivity} is phenomenologically consistent with \textsc{w-transparency} if, on it, we \textit{are} aware of our conscious states, yet it does not \textit{seem} like we are aware. Harman tells us “one might be aware of intrinsic features of experience without being aware of them \textit{as} intrinsic features of experience” (1990: 38). Thus a \textit{projectivist} could accommodate transparency (Schroer 2007).^59

---

^58 That said, the subjectivist theory of consciousness I develop in § 4.4 rejects a slightly different claim—viz. that awareness \textit{simpliciter} requires attention and/or availability for demonstrative reference. Tye (2010) ascribes to this view when it comes to \textit{visual} awareness (what Tye calls ‘phenomenal seeing’).

^59 Boghossian & Velleman’s (1989) projectivism is limited to colors, treating them as properties of the \textit{visual field}. For projectivist theories that treat phenomenal qualities as intrinsic properties of \textit{experience},
Mere phenomenological consistency is risky. After all, projectivism is an *error theory*. Unless we have strong reason to think otherwise, we should stick with an approach on which we can accommodate *W-*TRANSPARENCY without having to convict our experiences of massive error (cf. Tye 2002: 139). Parity seems to force us to place the same constraint on Higher-Order theory; all other things being equal, we should avoid precisifications of TRANSITIVITY that yield mere phenomenological consistency with *W-*TRANSPARENCY. In the present context, however, I propose we relax this constraint, but add the following proviso: it must be the case that it doesn’t seem like we are aware of our conscious states in both introspective and non-introspective awareness. That is, for this move to carry any force, TRANSITIVITY must be phenomenologically consistent with TRANSPARENCY\textsubscript{INT} too. The motivation for this rider is clear. Both non-introspective consciousness and introspective consciousness have concrete (although not necessarily different) phenomenological realities as part of their nature: there are facts of the matter concerning what it is like to be both non-introspectively and introspectively conscious. And because there are such facts of the matter, our theorizing can yield phenomenological predictions at odds with those facts. And that’s no less a problem in the introspective case than it is in the non-introspective case. If anything, it might be more of a problem; given the focused and often deliberate nature of introspective awareness, we might have good albeit defeasible reasons to think we get things right when we report what we find. Thus the rider is well justified, and the proponent of TRANSITIVITY ought to seek a precisification that allows for phenomenological consistency.

---

simpliciter. We’ll have the opportunity to look at a precisification with potential for a move along these lines in § 3.4.1.

There are three types of glosses. Each gloss trades on a distinction between the act of consciousness and the object of consciousness on HO theories. Act-Side Precisifications concern the state of higher-order awareness M*—the act of consciousness. Awareness comes in many forms, so an Act-Side Precisification will fill in the details. Object-Side Precisifications are ‘object-side’ because they concern the first-order state M itself—the object of consciousness. Are we aware of first-order mental state qua mental? Are we aware of merely M, or being in M? And so on. Finally, Complex Precisifications combine an Act-Side with an Object-Side Precisification, telling us about the act and the object of consciousness.

All precisifications can be seen as filling in the ‘in some way’ clause of TRANSITIVITY. So ‘in some way’ should be read broadly; it can be read as referring to the type of awareness relation (where ‘way’ modifies the act of consciousness), or it can be read as referring to the first-order mental state (where ‘way’ modifies the object of consciousness).

Consistency glosses are cheap. But a HO theorist cannot adopt just any consistency gloss. Consider two illustrative cases:

---

60 By calling M* an ‘act’ of consciousness, I simply mean that it is a state we have conscious awareness with, even if—as HO theorists would insist—it does not have the property being conscious, as it is not an object of consciousness (cf. Dretske 1997). While we can be aware of things in virtue of being in (just) M on HO theory, without M* we would not be consciously aware of anything.

61 NB: our precisifications are partial precisifications. The ‘in some way’ clause is unpacked only enough to comport with TRANSPARENCY. The HO theorist will ultimately want to add to these.
**Deflationary Transitivity:** Conscious mental states are mental states we are aware of automatically aware (in a way akin to how one automatically smiles one’s smile or dance’s one dance).\(^{62}\)

**Dispositional Transitivity:** Conscious mental states are mental states of which we are disposed to be aware.

Deflationary Transitivity and Dispositional Transitivity are both Act-Side Precisifications. On the former, awareness is automatic. In the same way laughing at a person necessarily involves laughing a laugh at that person, having an experience of something (i.e. undergoing a conscious state) necessarily involves experiencing an experience of something. On the latter, it is dispositional. ‘Aware’ plausibly admits a non-occurrent sense; we can be aware, as Daniel Dennett suggests, that “zebras in the wild don’t wear overcoats” merely if we are disposed to assent to that proposition in virtue of what we occurrently know (1978: 104).

On Deflationary Transitivity, all there is to the idea that in having an experience (or being in a conscious state), one is necessarily aware of it, is the triviality that in having an experience, one experiences one’s experience just as one smiles one’s smile or laugh’s one laugh. The problem is that this analogy is toothless. Laughing a laugh is just *laughing*; it’s not as if, when I laugh a laugh at a joke, I am laughing *at* my laugh (Tye 2009: 5). So, if experiencing an experience is like laughing a laugh, then experiencing an experience is just *experiencing* (cf. Rosenthal 2012c). This makes Deflationary Transitivity a consistency gloss—after all, W-Transparency doesn’t preclude our having experiences (or experiencing)—but it comes at a steep price. The point of invoking higher-

\(^{62}\) Sosa (2002), Block (2007) and Tye (2009) mention Deflationary Transitivity in passing, but there has been no sustained attempt to unpack it in any meaningful detail, save from the obvious truth that it is the sort of awareness that, if real, is reported by means of an internal accusative (Siewert 2012: 136). Indeed, neither Block nor Tye—the parties relevant here—officially endorse Deflationary Transitivity.
order awareness on HO theories is to distinguish between conscious and unconscious mental states, or (if you’ll grant the usage), conscious and unconscious experiences. **Deflationary Transitivity**, in effect, just says there are experiences, without telling us what makes something a conscious experience. No HO theorist can accept Deflationary Transitivity, since no HO theorist can accept that the sense in which we are aware of our conscious states merely comes down to this automatic, deflationary sense of awareness.

What about Dispositional Transitivity? A mental state might be available to a state of higher-order awareness despite our never actually (i.e. occurrently) being aware of it, just as a salt crystal might still be disposed to dissolve in water despite never getting wet. So W-Transparency isn’t violated. However, mere availability to higher-order awareness is a different beast. True, dispositional theories under a higher-order guise have been proposed (e.g. Carruthers 2000), but then again even Block—a staunch opponent of Higher-Order theories—doesn’t deny that lower-order mental states must be accessible to higher-order awareness, just that necessarily most are unaccessed (2011a: 567). In any event, I don’t want to quibble about whether a dispositional variant of HO theory should count as a ‘real’ HO theory. The point is that while there may be no tension between Dispositional Transitivity and W-Transparency, Dispositional Transitivity won’t cut it as a consistency gloss since it fails to motivate HO theory, at least as understood here.

The lesson from Deflationary and Dispositional Transitivity is a constraint on a consistency gloss. We alluded to the constraint in § 3.1. The constraint is this:

---

63 There is a wrinkle here. For this to be a consistency gloss, we need to say that something has a dispositional property F even though it is necessarily the case that F is never actualized. I grant that this might sound odd. But as Dispositional Transitivity is just meant to be an illustrative case of a precisification the HO theorist cannot accept, we can bracket this concern.
**THE COMPATIBILITY CONSTRAINT:** For any precisification P of TRANSITIVITY, if P is to be adopted as the ‘official’ precisification of TRANSITIVITY (so as to comport with W-TRANSPARENCY), P must be such that for at least one HO theory H, P is consistent with H.

H needn’t be an extant HO theory. H just need be a possible HO theory, where H is a HO theory if, according to H, a mental state is conscious only if it is the object of an occurrent, non-deflationary, non-inferential state of higher-order awareness.

This allows us to refine our thesis. Let’s say that a precisification P of TRANSITIVITY is an **acceptable** consistency gloss if (i) P is a consistency gloss (i.e. it comports with W-TRANSPARENCY); (ii) it entails TRANSITIVITY; and (iii) it meets THE COMPATIBILITY CONSTRAINT. So, my claim is that TRANSITIVITY admits of no clear acceptable consistency gloss. Clause (ii) is related to (iii) but it is not the same. If P does not entail TRANSITIVITY it will not meet THE COMPATIBILITY CONSTRAINT, but P can entail TRANSITIVITY and not meet THE COMPATIBILITY CONSTRAINT for some other reason.

My strategy will be to look at precisifications on a case-by-case base. I examine only the most promising options. I do not aspire to exhaustiveness. That is why I claim that TRANSITIVITY admits of no **clear** acceptable consistency gloss. But note that even this weaker claim (i.e. weaker than the claim that there is no such gloss) should suffice to help dislodge the notion with which we began, viz. that TRANSITIVITY is a truism. We’ll look at that issue again in § 3.5.64

Precisifications are divided into two groups. Group One (§ 3.4.1 - § 3.4.3) covers precisifications presented in the literature. I argue that these precisifications are not acceptable consistency glosses because they are not consistency glosses. Group Two (§ 3.4.4 - § 3.4.5) covers

---

64 So I won’t look at precisifications that appeal to acquaintance (as on HOA) or perception (as on HOP). If the appeal to perception isn’t idle, it is hard to see how (say) PERCEPTUAL TRANSITIVITY can help (cf. Rosenthal 2005). Likewise for ACQUAINTANCE TRANSITIVITY. (Acquaintance, after all, is de re and direct).
precisifications that, while not officially aired in the literature, take a more direct approach to W-TRANSPARENCY, say, by appeal to de dicto or indirect awareness. Though these precisifications are consistency glosses—there would be no reason to discuss them otherwise—they are not acceptable consistency glosses, as they do not meet THE COMPATIBILITY CONSTRAINT.

3.4 **THE MANY VARIETIES OF TRANSITIVITY**

3.4.1 **UNCONSCIOUS TRANSITIVITY**

Though not a HO theorist herself, Isabel Gois suggests that it is a confusion to think that HO theories have a problem with W-TRANSPARENCY. She says:

[The] criticism [that there is a conflict with W-TRANSPARENCY] misses the point of the [Higher-Order] strategy...[I]f a first order state is made conscious by a second order state, that second order state will not be conscious unless it is in turn made the object of a third order state...This explains [why] our experiences should seem transparent—we are typically never aware of the higher order mental state that makes a mental state conscious (2010: 148).

On its face, this move is puzzling. W-TRANSPARENCY is a claim about our (lack of) awareness of our conscious mental states. According to HO theory, conscious mental states are appropriately targeted *first-order* mental states. So that we are “typically never aware” of our higher-order states is neither here nor there (cf. Textor 2015: 2, fn. 5).

However, my guess is that what Gois actually has in mind is something bit subtler. Rosenthal often points out that experimental work on masked priming and blindsight in individuals with damage in their primary visual cortex reveals that perception can be subliminal (see, e.g. 2008: 837). This suggests a distinction between awareness and conscious awareness: subliminal perception makes us aware of things, but such states are conscious only when represented by a
HOT. Only then is our awareness of the world conscious. But our awareness of our first-order mental states is not itself conscious until we introspect, until we have—at least on the HOT-theoretic picture—a second HOT directed at the first HOT.

So maybe what Gois is after—what we can call ‘Rosenthal-Gois move’—is to point out that, while we are always aware of our conscious states, since that awareness is normally unconscious, it does not seem like we are aware of our conscious states, and hence why we find transparency compelling. This yields the following Act-Side Precisification:

**UNCONSCIOUS TRANSITIVITY:** Conscious mental states are mental states of which we are unconsciously aware.

**UNCONSCIOUS TRANSITIVITY** is incompatible with **TRANSPARENCY** since on **UNCONSCIOUS TRANSITIVITY** we are aware of our conscious states. So compatibility with **TRANSPARENCY** is an open question. Thus we’ll explore appeals to indirect awareness later, beginning in § 3.4.5. What remains for now is **TRANSPARENCY**, and it is here where the Rosenthal-Gois move runs aground. It may be that, in non-introspective consciousness, it does not seem like we are aware of our conscious states because that awareness is not conscious. Yet this doesn’t help if, when we do introspect, it does seem like we are aware of our conscious states. So **UNCONSCIOUS TRANSITIVITY** is phenomenologically consistent with non-introspective versions of **TRANSITIVITY**, but phenomenologically **inconsistent** with **TRANSPARENCY**. And given our earlier stipulation that the HO theorist cannot get with any phenomenological inconsistency, that’s a problem.

Indeed, Gois tacitly admits as much; she says our higher order states are only “typically” unconscious. This concession is damaging, since non-typical cases matter. For when we introspect, our higher-order awareness will be conscious by the HO theorist’s own lights, and so
our conscious states will be phenomenologically manifest, contra Transparency\textsubscript{INT}. Rosenthal himself comes close to admitting this:

Harman’s insistence that the only quality one is aware of...is that of the tomato...is too quick. When one sees a tomato consciously but unreflectively, one conceptualizes the quality one is aware of as a property of the tomato.

But when we introspect:

One thinks of the quality differently when one’s attention has shifted from the tomato to one’s experience of it. One then reconceptualizes the quality one is aware of as property of the experience; one then becomes conscious of that quality as the qualitative aspect of an experience, in virtue of which that experience represents a red tomato (2005: 120-121).

Now it might be possible for the HO theorist to go another route: she might simply say that that the higher-order awareness is necessarily unconscious. Here we adopt Unconscious Transitivity with a modification:

**Unconscious Transitivity*: Conscious mental states are mental states of which we are, necessarily, unconsciously aware.

Unconscious Transitivity* allows for phenomenological consistency. It is true that HO theorists disagree over whether our awareness of our mental states is normally conscious (Kriegel) or unconscious (Rosenthal). Nevertheless no HO theorist says that such awareness is necessarily unconscious. They can’t. Recall the story on Rosenthal’s HOT theory. Here, what it takes to be consciously aware of a mental state M rather than unconsciously aware is our having a state of higher-order awareness M* (directed at M) that is itself the object of yet another, distinct, state of higher-order awareness M**. So if Unconscious Transitivity* is true, there must be some reason we cannot instantiate M**, despite the fact that we can instantiate M*, and despite the fact that they are *the very same kind of state*. Perhaps there are limitations on the number of
levels of awareness normal humans can possess, but absent empirical evidence, there is no reason to think that the move from M* to M** is too cognitively burdensome.

In short, to adopt **UNCONSCIOUS TRANSITIVITY** would, in effect, be tantamount to adopting eliminativism about introspection.65 True, this would not be a violation of the Compatibility Constraint. But it is too implausible to take seriously, which is likely also why no one does.

In any event, I propose we bracket **UNCONSCIOUS TRANSITIVITY**. I also propose we bracket its even less plausible cousin, where all and only those first-order mental states that are transparent are the objects of a necessarily unconscious state of higher-order awareness, as entirely ad hoc. A better option, at least insofar **TRANSPARENCY** goes, is to bracket unconscious higher-order awareness, and to argue that what we are aware of must outstrip what we can introspect. That possibility is taken up next.

### 3.4.2 Peripheral Transitivity

Kriegel (2009) offers the following precisification of Transitivity:

**Peripheral Transitivity:** Conscious mental states are mental states of which we are peripherally aware.

Peripheral Transitivity is an Act-Side Precisification. Peripheral awareness contrasts with focal awareness, where the distinction between “focal and peripheral awareness is primarily a distinction regarding the distribution of attention” (ibid: 180). Typically, this is a distinction we apply to outer awareness, i.e. awareness of things in the environment. Kriegel thinks it applies just as well to inner awareness, i.e. awareness of mental states. Let’s grant that claim.

---

65 Eliminativism about introspection would strictly speaking only follow on a Rosenthalian HO theory. For other options on which this is not the case (e.g. Kriegel’s HO theory), see next in § 3.4.2.
On the face of it, **PERIPHERAL TRANSITIVITY** doesn’t jibe with **TRANSPARENCY\textsubscript{DE RE}** and **TRANSPARENCY\textsubscript{DIRECT}**. But this won’t trouble Kriegel, since he ascribes only to **TRANSPARENCY\textsubscript{INT}**. Kriegel frames transparency this way: for any experience $e$ and any feature $F$, if $F$ is an introspectible feature of $e$, then $F$ is part of the first-order content of $e$ (*ibid*: 181). A first-order content is a content involving neither representations nor representational properties. Higher-order content, recall, involves representations or representational properties. Since conscious states are representations (or at least Kriegel thinks so), he just means that conscious states are not introspectable. So his take is equivalent to the unrestricted version of our **TRANSPARENCY\textsubscript{INT}**.

Kriegel argues that **PERIPHERAL TRANSITIVITY** is consistent with **TRANSPARENCY\textsubscript{INT}** by appealing to an *attention-shift* model of introspection (*ibid*: 183). On this view, instead of involving a new representational state, introspection reorganizes the center/periphery structure of one’s overall experience. Suppose I have an experience of an apple. According to Kriegel’s self-representationalism, in non-introspective cases of consciousness, I have focal awareness of my apple but peripheral awareness of my experience of the apple. When I introspect, however, I am focally aware of my experience of the apple and peripherally aware of the apple (*ibid*).

How does this help square **PERIPHERAL TRANSITIVITY** with **TRANSPARENCY\textsubscript{INT}**? The idea is that when we introspect our experiences—i.e. stand in a relation of focal awareness to them—our prior peripheral awareness is “annihilated” (*ibid*: 184). We cannot introspect our experiences *qua* their being in the periphery if introspection is a matter of inner focal attention (*ibid*: 185). We can be aware of our experiences, then, without those experiences being introspectable. As such, there is no conflict between **PERIPHERAL TRANSITIVITY** and **TRANSPARENCY\textsubscript{INT}**. Moreover, since we are treating

---

\textsuperscript{66} But again, whether an appeal to de dicto or indirect awareness can help will be explored below.
W-TRANSPARENCY as a disjunction with three disjuncts, this is sufficient to show that there is no
conflict between PERIPHERAL TRANSITIVITY and W-TRANSPARENCY too.

I want to make two points in response. First, PERIPHERAL TRANSITIVITY is not a consistency gloss
without the attention-shift model of introspection. Yet Kriegel musters little evidence for this
view. The model supposedly has the “advantage of illuminating the fact that introspection does
not feel, phenomenologically speaking, like performing a ‘dramatic’ mental act, an act that
creates an altogether new representation” (ibid: 183). For my part, if introspection feels like
anything, it does feel like a ‘dramatic’ mental act insofar as introspection is effortful.

Granted, this is just a clash of intuitions. There is a deeper point, however. Even if we
conjoin PERIPHERAL TRANSITIVITY with an attention-shift model, and assume that the attention-shift
model is correct, pace Kriegel, we still don’t have a consistency gloss. Here’s why. Subjects are
aware of things. Conscious states, of course, aren’t. The idea is that subjects are made aware of
things by being in conscious states. Now on Kriegel’s picture, conscious states represent
themselves and represent the ambient environment. So subjects actually stand in two awareness
relations when in a conscious state; they stand in one to the conscious state itself (in virtue of
that state representing itself in the right way), and one to some environmental entity (in virtue
of that state also representing the environment). When we are non-introspectively conscious, we
stand in a relation of peripheral awareness to our conscious state, and a relation of focal
awareness to that environmental entity. Now what is certainly true is that if introspection is a
matter of standing in a relation of focal awareness to our conscious states, we are no longer
peripherally aware of our conscious states when we introspect. Peripheral inner awareness is
indeed annihilated. But what also needs to be annihilated is our standing in any introspective

TRANSITIVITY AND TRANSPARENCY
relation to our conscious state. In other words, it needs to be true that, when we introspect, all we are aware of is, in Kriegel’s terms, first-order content (ibid: 185). But this doesn’t follow. What follows is I am no longer peripherally aware of my conscious state, not that what I am aware of is limited to first-order contents.

And it is not clear why the fact (if it is a fact) that we are employing a shift in attention, rather than an entirely new representational state, should matter. When I introspect, my visual experience of a red apple represents both the red apple and itself, but does so with a certain characteristic distribution of attention. Instead of most of my attention being dedicated to the red apple (as in non-introspective consciousness), most of my attention is dedicated to my seeing the red apple. My experience of seeing the apple qua object of peripheral awareness is not introspectable. But that does not make Peripheral Transitivity consistent with Transparency\textsubscript{INT}, since it is consistent with this possibility that my experience is still an object of introspection—i.e. an object of focal inner awareness.\footnote{Kriegel is likely motivated here by showing how a bit of phenomenology—i.e. the phenomenology associated with peripheral inner awareness (his treatment of ‘subjective character’)—is consistent with Transparency\textsubscript{INT}. (For Kriegel, contra Rosenthal, the higher-order awareness is conscious). Given this, there is a sense in which his story is successful; you cannot introspect this bit of phenomenology, given that introspection is a matter of focal inner awareness. Nonetheless, it remains the case that on his view we can introspect higher-order content, and this is so precisely because he allows that we can have focal inner awareness of our experiences. (What else would the “reorganize” of our distribution of attention amount to if this weren’t so?). And this is enough to undercut any compatibility with Transparency\textsubscript{INT}.}

3.4.3 **Non-State-Relational Transitivity**

As it stands, Transitivity posits a relation between a higher-order state and a first-order state. But not all HO theorists construe Transitivity this way. Instead, some advance:
**Non-State-Relational Transitivity:** A conscious mental state is a mental state whose subject is aware of itself as being in the state.\(^{68}\)

**Non-State-Relational Transitivity** is an Object-Side Precisification. It seems consistent with each disjunct of **Transparency**. For—at least on the surface—it doesn’t even say that we are aware of our conscious states. It says that we are aware of *ourselves* as being in conscious states.\(^{69}\)

However, if it follows from **Non-State-Relational Transitivity** that we are unaware of our conscious states full stop, **Non-State-Relational Transitivity** fails the **Compatibility Constraint** for the simple reason that it entails that **Transitivity** is false. But if **Non-State-Relational Transitivity** fails the **Compatibility Constraint**, it is not an acceptable consistency gloss.

Here’s an argument for the claim that **Non-State-Relational Transitivity** entails that **Transitivity** is false, and so fails the **Compatibility Constraint**. Awareness is a relation. What we are aware of is what we stand in a (suitable) relation to. Yet on **Non-State-Relational Transitivity** we don’t stand in a relation to our mental states. As such, on **Non-State-Relational Transitivity** we are not aware of our mental states.\(^{70}\)

---

\(^{68}\) See, e.g. Berger (2014) and Brown (2012b, 2015), and fn. 11. I have added ‘state’ to Brown’s original ‘non-relational transitivity’ for the very reason he notes (2012b), viz. that calling this ‘non-relational’ *simpliciter* is misleading since we are still positing a relation between the representation and oneself, although there is no further relation between the higher-order state and some first-order state.

\(^{69}\) The motivation for **Non-State-Relational Transitivity** stems (in part) from its ability to address the objection from **radical misrepresentation**, i.e. the possibility of having a **HOT** that represents a state that one is not in (Neander 1998; Block 2011b). Briefly, here is how the move works. Distinguish between **weak factivity** and **strong factivity**. On weak factivity, if I am aware of \(x\), \(x\) exists. On strong factivity, if I am aware of \(x\) as being \(F\), not only does \(x\) exist, but \(x\) is also \(F\). Proponents of Non-State-Relational Transitivity reject **strong factivity**, accepting only **weak factivity**. In representing oneself as being in a certain state \(F\), I am guaranteed to exist, but I’m not similarly guaranteed to be in \(F\) (Rosenthal, private communication). If a mental state’s being conscious is just a matter of being suitably aware of ourselves as being in that state, that there is no first-order state in radical misrepresentation is otiose.

\(^{70}\) See Coleman (2015) for an independent argument that tallies well with this claim.
Does this argument work? I could see one rejecting the claim that awareness is a relation; adverbialists do just this. (Adverbialists can allow that we stand in a relation in virtue of being aware of something. What the adverbialist must reject is the converse, viz. that we are aware in virtue of bearing relations. For more on adverbialism, see § 4.3). Though no HO theorist actually does this, there is no in-principle reason why she cannot adopt adverbialism.\(^\text{71}\)

That said, I propose we just grant that NON-STATE-RELATIONAL TRANSITIVITY entails TRANSITIVITY, and push forward to other precisifications. The reason is that, as it stands, NON-STATE-RELATIONAL TRANSITIVITY is not a consistency gloss. It may well be, but not without supplementary claims concerning the way we are aware of our mental states. Absent that, NON-STATE-RELATIONAL TRANSITIVITY cannot help the HO theorist worried about W-TRANSPARENCY.

3.4.4 De Dicto Transitivity

Here is where we are. The precisifications we have encountered—DEFLATIONARY, DISPOSITIONAL, UNCONSCIOUS, PERIPHERAL, and NON-STATE-RELATIONAL TRANSITIVITY—have all been aired in the literature, albeit not always with TRANSPARENCY in mind. None of them worked out. DEFLATIONARY and DISPOSITIONAL TRANSITIVITY were consistency glosses but not acceptable consistency glosses, while PERIPHERAL, UNCONSCIOUS and NON-STATE-RELATIONAL were not even consistency glosses.

At this juncture a more direct approach may prove fruitful. W-TRANSPARENCY is a disjunction with three disjuncts. The HO theorist is thus free to simply pick one disjunct and aim for consistency with that disjunct since (of course) consistency with one disjunct is sufficient for consistency with W-TRANSPARENCY. Suppose we opt for TRANSPARENCY\(_{DE.RE.}\). Then we could say:

---

\(^{71}\) Although Kriegel (2011) flirts with, but falls short of endorsing, this option, albeit for different reasons.
**DE DICTO TRANSITIVITY:** Conscious mental states are mental states of which we have fact-awareness.

**DE DICTO TRANSITIVITY** is an Object-Side Precisification. Dretske (1993, 1995, 1999) famously distinguished between fact-awareness (or ‘awareness of facts’) and thing-awareness (or ‘awareness of things’). Awareness of things comes in three variants: awareness of objects, properties, and events. For simplicity, I’ll speak generically of awareness of things. Unlike awareness of things, awareness of facts implies a deployment of concepts. If I am aware that the water is boiling, then I must possess, and employ, the concept WATER and BOILING in my awareness of the water (1993).

I am going to assume that fact-awareness and thing-awareness are indeed distinct; the reports ‘I see the blue Ming vase’ and ‘I see that the Ming vase is blue,’ on this view, describe two distinct mental states. I’ll assume this, but it is worth noting—if only briefly—that there are strong reasons in its favor. A key purported difference between the two states is, as noted, a matter of concept-deployment. Yet perhaps that’s wrong. Giustina and Kriegel (forthcoming) suggest that perhaps there must be at least some fact you are aware of regarding the thing you see. Maybe you’re not aware of what you see as a Ming vase. Still, you might be aware of it as, say, an entity. If that is so, the distinction crumbles; there is always some concept—namely the concept ENTITY—that is always employed when a subject is aware of something (forthcoming).

However, as Giustina and Kriegel (forthcoming) point out, there reasons to resist this. First, whether such a maximally generic concept exists is controversial. Giustina and Kriegel rightly note that concepts partition entities: a concept F distinguishes the Fs from the non-Fs. A concept like

---

72 Dretske’s distinction concerns (outer-directed) perceptual awareness, but as Giustina & Kriegel (forthcoming) make clear, this distinction applies equally to other forms of awareness.
ENTITY that applies to literally everything could partition nothing. Secondly, it is doubtful that non-human animals and infants possess such abstract concepts, yet it is far less doubtful that they lack awareness (forthcoming).\textsuperscript{73}

On top of this, as Giustina and Kriegel (forthcoming) also note, unlike reports of thing-awareness, reports of fact-awareness are referentially opaque. We can infer from Carmichael’s seeing the third Ming vase from the left, and from the fact that the third Ming vase from the left is the most expensive, that Carmichael sees the most expensive Ming vase. However, we cannot infer from Carmichael’s seeing that this Ming vase is the third from the left, and from the fact that the third Ming vase from the left is the most expensive, that Carmichael sees that the Ming vase is the most expensive.

These points are not beyond dispute. But given that the distinction between thing and fact-awareness is widely accepted, I’ll adopt it going forward. Now, ordinary talk of our being aware of our conscious states is, like our talk about awareness of things in the world, neutral on whether such awareness is de re. The sentence ‘I am aware of my desire for tiramisu’ can be given a non-de re reading, viz. ‘I am aware that I have a desire for tiramisu’ (Tye 2014). DE DICTO TRANSITIVITY picks up on this insight. TRANSPARENCY\textsubscript{de re} permits fact-awareness, so DE DICTO TRANSITIVITY is a consistency gloss. The problem is that it doesn’t entail TRANSITIVITY, and so doesn’t meet THE COMPATIBILITY CONSTRAINT.

To see why, note that while we assumed that thing-awareness does not entail fact-awareness, it is also plausible that even if fact-awareness is based on some thing-awareness, one might be aware that something is F without being aware of that thing that is F. So, strictly

\textsuperscript{73} As Giustina & Kriegel also point out, Dretske (1993: 268) is well aware of these objections.
speaking, my being aware that I am in a conscious state does not entail I am aware of my conscious state. Take Dretske's example: one can be aware of the fact that the water is boiling (aware that it is boiling) without being aware of the objects or the properties that constitute the fact that the water is boiling (1999). I may hear a timer go off, but if I am looking the other way, I am not (de re) aware of the water, the flame, or the property of being boiled. Fact-awareness thus splits off from constituent thing-awareness when the former is indirect. Fact-awareness is indirect when it is achieved by awareness of properties and/or objects other than those involved in the fact, and direct otherwise. Awareness that your child has a temperature is indirect when you use a thermometer but direct when you feel his forehead (ibid: 108).

The HO theorist might try to amend De Dicto Transitivity by stipulating that the fact-awareness is direct. This would yield the following complex precisification:

**Direct De Dicto Transitivity:** Conscious mental states are mental states of which we are directly fact-aware.

Since Direct De Dicto Transitivity entails Transitivity the aforementioned problem is rectified. Yet it is a pyrrhic victory, for Direct De Dicto Transitivity is not a consistency gloss. It is not a consistency gloss because it says we do have de re awareness of our conscious states. Indeed, this is exactly what proponents of transparency often warn against. Here is Tye:

[W]e can be aware that we are having such experiences. No one denies that. But such factive awareness is not supported by awareness of, or attention to, the token experiences themselves. This is the familiar and widely accepted doctrine of transparency (2009: 5).

So De Dicto Transitivity and Direct De Dicto Transitivity seem to fail. The former fails because it does not entail Transitivity, and the latter because it is not a consistency gloss.

What this shows, I take it, is that the HO theorist will have to reject Transparency De Re outright; if there is hope for an acceptable consistency gloss, it will be had by going in for
and adopting a ‘transparency-friendly’ account of introspection, or by sticking with de re awareness, but insisting that such awareness is indirect, opting for TransparencyDirect. We have already seen why the former move likely fails. We’ll broach the latter next.

3.4.5 INDIRECT TRANSITIVITY

All that remains for the HO theorist, then, is TransparencyDirect.74 This yields:

INDIRECT TRANSITIVITY: Conscious mental states are mental states of which we are indirectly aware.

INDIRECT TRANSITIVITY is an Act-Side Precisification. It tells us that the relation of awareness is indirect. Since TransparencyDirect rules out only direct awareness, invoking indirect awareness is fine. The question is whether INDIRECT TRANSITIVITY meets THE COMPATIBILITY CONSTRAINT.

We discussed cases of direct and indirect fact-awareness earlier. But we haven’t yet given an explicit definition of direct and indirect awareness. Let’s do that now. Here I quote Stoljar: “A subject S is indirectly aware of x just in case S is aware of x by being aware of y, where y is distinct from x,” and “S is directly aware of x just in case (i) S is aware of x and (ii) S is not (merely) indirectly aware of x” (2004: 380).75 The indirect object on INDIRECT TRANSITIVITY will just be the conscious state. Yet there must also be something that anchors our being indirectly aware of our conscious state; there must be a direct object of awareness. The obvious option—indeed the option proponents of TransparencyDirect like Tye have pushed—is that the anchor is whatever that

74 The other option, TransparencyInt, only works if we can be aware of something that is in principle, introspectively inaccessible. We explored the viability of this in our discussion of Peripheral Transitivity.
75 Stoljar is following the work of William Alston on epistemic immediacy (1971) and Frank Jackson on perceptual immediacy (1977). The ‘merely’, again following Stoljar (ibid), is required to rule out the possibility of being both directly and indirectly aware of something (e.g. being aware of the coffee cup by feeling its heat, and seeing it).
conscious state represents. So the emerging picture would now be this. Suppose Carmichael has a first-order visual perception of a blue sky. By TRANSITIVITY, if that perception is conscious, Carmichael instantiates a state of higher-order awareness M* that has as one of its objects the first-order visual perception. INDIRECT TRANSITIVITY would then add to TRANSITIVITY the following qualification: Carmichael is conscious of his visual perception by being conscious of what that perception represents, viz. a blue sky.

I have never seen a HO theorist advance anything like INDIRECT TRANSITIVITY. This is for good reason in my view, but this does not tell against its meeting THE COMPATIBILITY CONSTRAINT. After all, INDIRECT TRANSITIVITY still entails TRANSITIVITY, so we can still say that conscious mental states are mental states we are aware of, albeit indirectly. So INDIRECT TRANSITIVITY deserves a closer look. One question we might ask is whether we are aware of a fact (‘indirect-fact awareness’) or a thing (‘indirect-thing awareness’). This yields two more complex precisifications:

**INDIRECT-FACT TRANSITIVITY:** Conscious mental states are mental states of which we have indirect-fact awareness.\(^{78}\)

**INDIRECT-THING TRANSITIVITY:** Conscious mental states are mental states of which we have indirect-thing awareness.

We saw one reason why INDIRECT-FACT TRANSITIVITY is out. Like DE DICTO TRANSITIVITY, INDIRECT-FACT TRANSITIVITY does not entail TRANSITIVITY, since awareness *that* I am in a conscious state does not entail awareness *of* a conscious state. Therefore, it is not an acceptable consistency gloss.

---


\(^{77}\) FO theorists like Tye and Dretske would concur; on FO theories, this is just how we introspect our conscious states, not what makes a mental state conscious.

\(^{78}\) Again, the ‘of’ in INDIRECT-FACT TRANSITIVITY marks a grammatical compromise, not de re awareness.
Here is another potential reason. **INDIRECT-FACT TRANSITIVITY** is based on Dretske’s displaced model of introspection. By Dretske’s lights, this makes the higher-order awareness in **INDIRECT-FACT TRANSITIVITY** *inferential* (1995: 42). The idea is this. Suppose I am aware that an experience is F. Then there has to be a premise connecting our perceptual *de re* beliefs about the environment (e.g. based on my seeing that my experience’s object has some property $F^*$) to our introspective belief that we are having an experience with F (e.g. that the object would not have $F^*$ unless that experience had F). Yet HOT, HOP, Self-Representational, and Higher-Order Acquaintance theories all deny, rightly, that the relevant higher-order awareness is inferential. One can infer from publicly accessible evidence that one is angry without the anger being thereby rendered conscious. Thus the higher-order awareness we have of our anger must be non-inferential for that anger to be conscious *(Rosenthal 1997b: 737)*.79

**INDIRECT-THING TRANSITIVITY**, by contrast, fares better. Awareness that I am in a conscious state (fact-awareness), we saw, does not entail awareness of my conscious state. But **INDIRECT-THING TRANSITIVITY** does not implicate fact-awareness. It implicates thing-awareness. So unlike **INDIRECT-FACT TRANSITIVITY**, **INDIRECT-THING TRANSITIVITY** entails **TRANSITIVITY**. The question, then, is whether an appeal to *indirect* awareness scuttles the **COMPATIBILITY CONSTRAINT**. We are indirectly aware of y only if we are aware of y by being directly aware of x. The key is how we understand the term ‘by.’ ‘By,’ and similar expressions like ‘based on,’ do not wear a unique interpretation on their face *(Alston 1971)*. We just encountered one interpretation that won’t cut it: to say that I am (indirectly) aware of y by being (directly) aware of x is to say that I *infer* that y from x.

79 That inference is required is not uncontroversial. Tye denies the need for inference (2002: 144). Yet as Murat Aydede points out (2003: 62, fn. 5), Tye just flatly asserts that inference is not needed.
So the proponent of INDIRECT-THING TRANSITIVITY will need something else. Another option is
to gloss the ‘by’ with the *in virtue of*-relation: to say that I am aware of *y* by being aware of *x* is
to say that I am aware of *y* *in virtue of* being aware of *x*. To see what this amounts to, we can start with a more mundane example, adapted from Frank Jackson’s (1977). Suppose that when Carmichael parks his car on a busy street, his car makes actual physical contact with curb. It is natural to say that Carmichael’s car touches the curb *in virtue of* its tire touching the curb. William Fish (2004) offers a nice way of cashing this out. We have two states of affairs: a *tire-touching-curb* state of affairs and a *car-touching-curb* state of affairs. The latter clearly depends on the former: Carmichael’s car will come into contact with the curb only when a part of the car does (*ibid*: 6). But the converse does not hold: whether the former state of affairs constitutes the latter depends, as Fish notes, “upon the other relationships—relationships that do not alter the underlying *kind* of state of affairs it is—it enters into” (*ibid*). The *tire-touching-curb* state of affairs is a type of state of affairs that could have obtained even if it fails to constitute a *car-touching-curb* state of affairs. This suggests, to quote Fish, the following take on ‘in virtue of:’ “an *S*’s φ’ing is in virtue of *p* only if *p* could have obtained otherwise, and *p*’s obtaining would not thereby have constituted *S*’s φ’ing” (*ibid*: 7).

Applying Fish’s account on the ‘in virtue of’ relation to our construal of indirect awareness, the pertinent question becomes this: Can the HO theorist say that Carmichael is (indirectly) aware of his conscious state in virtue of being (directly) conscious of what that state represents, in the sense that Carmichael could have been conscious of what that state represents *without* his thereby being aware of his conscious state? No. The problem is that this reverses the HO theorist’s order of explanation. On HO theory, the *Carmichael-conscious-of-blue sky* state of
affairs is not the kind of state of affairs that can obtain independently of Carmichael being aware of his visual sensation of the blue sky. Subliminal perception might make me aware of things in my environment, but that awareness is not conscious without higher-order awareness of my perceptual state.

I thus submit that INDIRECT-FACT TRANSITIVITY fails to meet THE COMPATIBILITY CONSTRAINT. Consequently, INDIRECT-FACT TRANSITIVITY is not an acceptable consistency gloss. Indeed, absent a more amenable way of cashing out the ‘by’ in indirect awareness, this point generalizes: any precisification of TRANSITIVITY that appeals to indirect awareness is not a consistency gloss. Thus we get perhaps our strongest claim yet: TRANSITIVITY entails that TRANSPARENCY_{DIRECT} is false.

3.5 **NOT SO TRUISTIC AFTER ALL?**

I have argued that there is no readily available precisification of TRANSITIVITY’s ‘in some way’ clause that is compatible with the various ways of precisifying W-TRANSPARENCY’s ‘in some way’ clause, or that least none that continues to motivate a HO theory. What now?

I think the appropriate reaction is to rethink the role of both TRANSITIVITY and W-TRANSPARENCY in our theorizing about consciousness. This means rethinking their status as truisms. To say that TRANSITIVITY and W-TRANSPARENCY are not truisms about consciousness is not to say that either thesis is false, or that either thesis is wholly unintuitive. We do sometimes cease to treat a proposition $p$ as a truism because we find out that that $p$ is false, but it also seems reasonable to cease bearing this attitude towards $p$ when we realize that the truth of $p$ entails that another widely held distinct proposition $p'$ is false, even though we are not sure which side to take. And this often cuts both ways; it might be the case that we should cease treating $p'$ as a truism as
well. There might be an epistemic symmetry between the attitudes we should take towards $p$ and $p'$ when $p$ and $p'$ are in conflict. So, to say that Transitivity and W-Transparency are not truisms can also imply that we should simply update—in this case, decrease—our degree of confidence or credence in both Transitivity and W-Transparency. And that, I venture, is likely what is going on here.

How does this play out in practice? One thing we can say for sure is that if Transitivity and W-Transparency are not truisms, they ought not function in arguments as if they are truisms—they ought not, in other words, function as innocuous assumptions in need of no defense. This is especially relevant for Transitivity, whose truth rules out an entire family of theoretical approaches to consciousness. To be sure, I am not claiming that Transitivity always functions in arguments as if it is a truism, or that no arguments for Transitivity have been aired. But the point is that Transitivity often functions in this matter, and that its functioning in this manner is only justifiable if it is genuinely a truism. If Transitivity cannot be rendered compatible with W-Transparency, not only is Transitivity not a truism, but also there is simply no argument for Ho theory that does not require an argument against W-Transparency.

---

80 In an appendix to his (2009), Kriegel does a nice job canvassing several arguments for Transitivity.
81 A clear example of this is the so-called ‘Simple Argument’ (e.g. Lycan 2001; Kriegel 2006), which takes Transitivity as a self-evident thesis, in essence delivering Ho theory for free.
4 SELF-CONSCIOUSNESS DESPITE ELUSIVENESS

4.1 SUBJECTIVIST THEORIES OF CONSCIOUSNESS

Call a subjectivist theory of consciousness a theory according to which self-awareness is a constitutive feature of phenomenal consciousness. On weak subjectivism, awareness of one’s mental states—weak self-consciousness—is a constitutive feature of phenomenal consciousness (e.g. Brentano 1874/1973; Kriegel 2009). On strong subjectivism, awareness of oneself—strong self-consciousness—is a constitutive feature of phenomenal consciousness (e.g. Kriegel 2003b; Rosenthal 2005; Gennaro 2011; Sebastián 2012; Zahavi 2014). On unconscious subjectivism, it is unconscious self-awareness that is a constitutive feature of phenomenal consciousness (e.g. Rosenthal 2005; Gennaro 2011). On conscious subjectivism, it is conscious self-awareness that is a constitutive feature of phenomenal consciousness (e.g. Kriegel 2003c, 2009; Sebastián 2012; Zahavi 2014).

82 ‘Subjectivism’ comes from Billon & Kriegel (2015), who use it to pick out theories on which phenomenal consciousness includes a sense of “for-me-ness.” On my usage, subjectivist theories includes such views, but are not exhausted by them. I will not explicitly engage such views here.

83 It has been suggested that the distinction between weak and strong self-consciousness collapses if we are talking about token mental states; i.e. consciousness of token mental states implies consciousness of oneself (e.g. Rosenthal 1997a, although see Kriegel 2009 for critical discussion). Since our focus going forward will be restricted to strong self-consciousness, we can rest easy ignoring this complication.

84 Sebastián (2012) makes a distinction between “mental state-involving” and “self-involving” theories of consciousness. Sebastián’s distinction is semantic; it concerns two different ways of understanding the content of experience (e.g. on a self-involving view, experience is such that its correctness conditions concern the subject undergoing the experience [ibid: 161]). Ours is orthogonal to Sebastián’s; to claim that (say) strong self-consciousness is a constitutive feature of phenomenal consciousness is not necessarily to make a claim about content. One could be a subjectivist and deny that conscious states have intentionality at all. That said, the positive view put forth (§§ 4.4 – 4.6) does appeal to content.

85 For the purposes of allowing this form of subjectivism, I will follow Rosenthal (2005) and use ‘aware’ and ‘conscious’ (and their cognates) interchangeably, but not in a way that makes ‘conscious awareness’ redundant. That said, since I am not concerned with unconscious subjectivism in this chapter, one can feel free to bracket unconscious self-awareness if need be.
All weak subjectivists are HO theorists since all weak subjectivists ascribe to TRANSITIVITY. All HO theorists are weak subjectivists since all HO theorists ascribe to TRANSITIVITY. But one can be a strong subjectivist and be either a HO theorist or a FO theorist. One can be a HO strong subjectivist if one thinks that strong self-consciousness entails weak self-consciousness, or if one thinks that weak self-consciousness entails strong self-consciousness. Likewise, one can be a FO strong subjectivist if one thinks that strong self-consciousness is essentially and exhaustively world-directed (e.g. Smith 1986; Sebastián 2012; Zahavi 2014).

We can capture this way of cutting up the space of possibilities with the following table:

<table>
<thead>
<tr>
<th>Subjectivist Theories of Consciousness</th>
<th>Weak Self-Awareness</th>
<th>Strong Self-Awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Order Theory</td>
<td>N/A</td>
<td>Sebastián 2012; Zahavi 2014</td>
</tr>
</tbody>
</table>

86 At least all ambitious HO theorists ascribe to weak subjectivism, since subjectivism concerns phenomenal consciousness. In § 2 I assumed, with Block, that Lycan’s HOP theory is a non-ambitious form of HO theory (see fn. 12). But Lycan himself is not always clear. Lycan claims that HOP is not a theory of phenomenal character (2004: 97), yet insists that ‘what it is like’ for the subject is “completely explained” by inner sense (1996: 4). So erring on the side of completeness, I have included Lycan in our taxonomy. Nothing will turn on this if, in the end, Lycan would resist describing HOP theory as a species of subjectivism, preferring to stick with Block’s ‘non-ambitious’ label instead.

87 Kriegel originally ascribed to strong conscious subjectivism (2003b, 2003c). However, he now appears to ascribe to weak, conscious (self-representational) HO subjectivism:

[M]y current experience...strikes me as egological—that is a form of peripheral self-awareness...There is an elusive sense of self-presence or self-manifestation inherent in even a simple conscious experience of the blue sky. It is less clear to me, however, that this feature of peripheral inner awareness—its being self-awareness and not mere inner awareness—is constitutive of the phenomenology (2009: 177).

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
My aim in this chapter is to advance a novel form of strong conscious subjectivism that is orthogonal to the debate between FO and HO theories. I call this theory *Egocentric Representationalism*, or ‘ER’ for short. ER is designed to meet a phenomenological challenge unique to strong conscious subjectivism. The challenge, at least at first glance, is simple. Strong conscious subjectivism claims that strong self-awareness is a constitutive feature of phenomenal consciousness. Yet there is a storied and diverse tradition according to which the self is phenomenally elusive. The worry, then, is that strong conscious subjectivism makes a phenomenological prediction that does not check out; the phenomenological facts, at least as many see them, do not seem to jibe with what strong conscious subjectivism tells us.

I advance ER not because I believe in ER. Rather, I advance ER because I believe it might be the best hope for the subjectivist who accepts the elusiveness of the self. ER, I’ll contend, is well poised to grant that the self is phenomenally elusive yet deliver self-consciousness despite the self’s elusiveness. What follows is a theoretical exercise: I won’t argue for strong subjectivism, and I won’t argue that the self is indeed phenomenally elusive. I’ll just assume both claims, investigate what constraints the latter put on the former, and build a theory to meet those constraints. My thesis is that ER does the job, and does it exceedingly well.

---


89 So given our focus, I will largely drop the ‘strong’ and ‘conscious’ qualifiers; strong conscious subjectivism will just be ‘subjectivism’ and strong self-consciousness will just be ‘self-consciousness’, unless I make explicit otherwise. Also, although there will be contexts where retaining ‘phenomenal’ will be useful, I will largely dispense with this qualifier as well. Thus, unless I make explicit otherwise, it should be understood that by ‘consciousness’ I mean phenomenal consciousness.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
This chapter is divided into three parts. The thesis that the self is elusive has, as we said, a rich history. Alas, this rich history has not, to my estimation, brought much by way of clarity (cf. Howell 2010; Strawson 2013). The thesis remains obscure. So in part one (§ 4.2), I carefully articulate (what I take to be) a plausible reading of the thesis, set aside its nearby but implausible cousins, and in doing both these things draw out the resultant constraint on subjectivism. Part two (§ 4.3) critically examines ER’s chief rival, viz. the adverbial approach to subjectivism (e.g. Thomasson 2000; Rowlands 2013; Zahavi 1999, 2006b, 2014). Finally, part three (§§ 4.4 – 4.6) lays out ER. The core of ER is in line with the spirit of pre-reflective self-consciousness familiar from the phenomenological tradition. Developing on an insight from Bernard Lonergan, ER claims that self-consciousness consists in our representing ourselves as standing in an irreflexive relation to objects of consciousness.

4.2 THE CHALLENGE FROM ELUSIVENESS

4.2.1 ARTICULATING ELUSIVENESS

It is difficult, if not impossible, to talk about the elusiveness of the self without saying something about the nature of the self. Doing so presents a challenge in its own right, however, since the concept SELF is arguably a mongrel (Olson 1998).

90 I won’t focus on (strong) subjectivists that are not rivals, i.e. subjectivists that reject the thesis that the self is phenomenally elusive. For instance, Rosenthal’s HOT theory is a species of strong unconscious subjectivism. According to Rosenthal, a mental state’s being conscious must consist “in one’s having a suitable thought that one is, oneself, in that state” (1997a: 74), and that having such a thought (a HOT) will “thereby make one aware not only of that state, but also of a self the HOT represents the state as belonging to” (2012c: 25). Rosenthal still addresses elusiveness because on his view the self-awareness can be conscious. On his view, Humean elusiveness was predicated on the presupposition of a perceptual model of inner awareness (2004, 2012c). I’ll just note here that it is unclear to me that all it takes to show that the self is not elusive is to point to our ability to have de se thoughts; it’s hard to see how proponents of the elusiveness thesis would deny our having such ability. For more on Rosenthal’s version of subjectivism, see fn. 103.
The best bet in my view is simply to stipulate: a ‘self’, as I use the term, is a subject of experience. Whatever else selves are (e.g. loci of practical concern), they are fundamentally experiencers and thus subjects, i.e. those things, whatever they may be, that have experiences. I will also assume that selves exist, and that you, me, and my Uncle Carmichael are amongst the many things that have been, are, and (hopefully) will continue to be selves. My own view, with Galen Strawson (2009), is that anti-realism about the self, understood as anti-realism about subjects of experience, is a hard road, perhaps bordering on absurdity. I think this because I am realist about experiences. Experiences are always experiences for someone. We cannot think of token experiences absent a subject of experience (ibid).\(^91\)

That is all we’ll take on board. So our realism is sparse. That’s as it should be. We needn’t make any assumptions about the nature of subjects. We won’t assume selves are physical. ER can accommodate a naturalist and a non-naturalistic perspective on the self. We won’t even assume that mental events within a single life are always instantiated by the same subject.\(^92\)

With these metaphysical niceties behind us, we are ready to articulate what is meant by the idea that the self is elusive. So, officially:

**ELUSIVENESS**: From the first-person perspective, the self is not phenomenally manifest as an object of experience.

I unpack **ELUSIVENESS** in three steps. First: the self is elusive from the first-person perspective. For our purposes, I follow Robert Howell in understanding the first-person perspective as “the

\(^91\) On most readings of his ‘bundle theory,’ Hume would reject realism about selves precisely because the self was elusive. (Hume might identify the self with the bundle, but that is not how I understand the term ‘self’ here). However, Hume’s anti-realism comes with further empiricist assumptions, and so there is nothing about the claim that self is elusive as such that makes anti-realism compulsory.

\(^92\) For relevant discussion of such “strong logical subjects,” see Geoffrey Lee (Draft).
perspective that one can take on oneself that is not available to another”, which excludes “the perspective one has on oneself by looking in a mirror, viewing a CAT-scan, or touching one’s nose” (2010: 460). This also rules out some purported counterexamples to ELUSIVENESS as irrelevant—e.g. that other people are aware of me (Howell *ibid*).

In addition, awareness one has of one’s body is not a counterexample to ELUSIVENESS (Chisholm 1969), and even on views where we are *essentially* embodied (Peacocke 2014). No proponent of ELUSIVENESS denies that we are aware of our body.

Second: ELUSIVENESS says that the self is not *phenomenally manifest*. Commitment to an x’s being phenomenally manifest entails a commitment to realism about x-experiences. We are realists about ‘Ming vase Experiences’ because Ming vases are, or can be, phenomenally manifest; my experience has the phenomenal character it does because it is as of a Ming vase. We are also *anti*-realists about all sorts of experiences. Take ‘Monday experiences’ (Kriegel 2011: 44). We are anti-realists about Monday experiences because Mondays cannot be phenomenally manifest; even though some experiences have the property of occurring on a Monday, no experience has the phenomenal character it does because it occurs on a Monday, or because it is of a Monday—we cannot experience Mondays. To claim that the self is not phenomenally manifest, then, is to say that selves are more like Mondays than like Ming vases.

But ELUSIVENESS does not say that selves are not phenomenally manifest *simpliciter*; ELUSIVENESS says that the self is not phenomenally manifest *as an object of experience*. This brings us to our third point. An object of experience is something like the Ming vase; when Carmichael has a visual experience of a Ming vase, the object of his experience is the Ming vase. We can also speak of Carmichael having a visual experience of the *blueness* of a Ming vase. The same goes for SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
other specific sensory modalities: whether I am seeing a Ming vase, smelling garlic or hearing a bird’s humming, something is manifest to me as an object.

These cases are standard, but my use of ‘object’ is somewhat technical. For one thing, it’s metaphysically neutral: the vase is a particular concrete entity, but its blueness is a property. This fits with an innocuous and perhaps more natural use where ‘object’ just means what we experience. It is true that in each case above—Carmichael having a visual experience of blueness, Carmichael having an olfactory experience of garlic, and so on—the object of experience is indeed what Carmichael experiences. But I mean something more than this. To say that \( x \) is phenomenally manifest as an object is to say that \( x \) is available for demonstrative reference (cf. Valberg 1992: 7), and it is so available precisely because of the phenomenology typified by \( x \)-experiences. That Carmichael can demonstratively refer to the Ming vase is grounded in the fact that things seem, phenomenally speaking, as if there is a Ming vase on the mantle (cf. Gertler 2012). I will call experiences like this—experiences where what I am experiencing is available for demonstrative reference—instances of object-consciousness.93

Object-consciousness is not limited to perceptual experience. It includes introspective experience as well. In the perceptual case, I might refer to that Ming vase on the mantle (an object), or that color of a Ming Vase (a property), or that falling of a Ming vase off the mantle (an event). Yet as Brie Gertler (2012: 106) makes clear, this holds for the introspective case too; I might refer to that reddish sense datum (an inner mental object, supposing there are any), this pinching feeling (a property), that experience of seeing a reddish sense datum (an event).94

93 So to say that something is given as an object of experience is not, at least here, to say that what is given must be conceptualized as an object. It is just to point to a phenomenological fact.
94 A few notes here. First, I assume, following Gertler, that whether a demonstrative refers to a particular,
Now there is a tempting line of reasoning that might seem to undercut ELUSIVENESS from the get go, but in reality rests on a misunderstanding that’s worth drawing out. The line starts with a fairly banal observation: we perceive things by perceiving their properties (e.g. Shoemaker 1996: 10). When Carmichael perceives the Ming vase, he does so by perceiving its various properties, like its color, size, and shape. But if that’s right, then why can we not say something similar about self-consciousness: worries about transparency (§ 3.2) and an inner sense model of introspection aside, if I can perceive (or in some other sense be aware) of my mental states, and my mental states are properties of myself, then it would seem that ELUSIVENESS is just false. I am aware of myself by being aware of my mental states.95

Here’s the misunderstanding, though. The proponent of ELUSIVENESS will grant (or at least can grant) that mental states are properties of selves. And she will grant (or at least can grant) that we perceive things by perceiving their properties. But she will deny that the metaphysical fact that these mental states are my properties will, as Howell puts it, be “reflected in phenomenology” (2010: 470). The proponent of ELUSIVENESS can grant that we are (de facto) aware of ourselves yet point out that it nonetheless never seems this way, and that the fact that

property, or event will depend (in part) on the subject’s referential intentions, and that this can hold for both the perceptual and introspective case (2012). Second, this neutral use of ‘object’ also allows me to accommodate views on which intentional relations to properties individuate experiences (Dretske 1995; Johnston 2004). Third, insofar as introspective demonstratives in particular go, even if it is true that the introspective ‘that’ should not be construed in terms of something like the public linguistic ‘that’ (Howell 2007), exactly how introspective demonstratives work won’t matter. All the proponent of ELUSIVENESS requires is that there is something like an introspective demonstrative form of awareness, however it is analyzed aside. For a particular account of introspective demonstratives (in terms of acquaintance), see Gertler (2012). Finally, by characterizing object-consciousness in terms of availability for demonstrative reference, I only mean to pick out a necessary condition on what it is for something to be an object of experience. Part of my positive aim, in fleshing out ER, will be to say a bit more about what being an object of experience consists in.

95 Roderick Chisholm (1969) argues in this manner, and Bertrand Russell (1912: 79-80) makes a similar move, albeit by appeal to acquaintance and sense data.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
it never seems that way is precisely the point. Howell explains this by having us compare self-awareness with a more basic case: Carmichael’s perceiving the Ming vase by perceiving its roundness. Here, the vase’s properties appear as properties of a ‘that,’ while the self’s properties do not (ibid: 471). Yet this difference, as Howell notes, is key. When Carmichael looks at the Ming vase and says ‘That looks beautiful’ he need not be pointing to a property: the apple itself is phenomenologically manifest as an object, and thus available for demonstrative reference. By contrast, when one introspects, even supposing that what is available for demonstration are properties of the self, there is no “co-locational” or “presentational unity” that (unlike the apple) allows us to say ‘that’ and refer to the self (ibid: 471). In sum: ELUSIVENESS is not (or need not) be the stronger claim that being (de facto) aware of something by being aware of its properties breaks down in the case of self-awareness. Rather, it is only the weaker (but perhaps more interesting) claim that this awareness is never manifested phenomenologically in object-consciousness.

4.2.2 MOTIVATING ELUSIVENESS

ELUSIVENESS should be surprising. We might wonder: why it is that a Ming vase can be phenomenologically manifest as an object of experience, but the self cannot? Ontologically

96 Colin McGinn makes something like this point in the following way:

When I think of myself that which thinks occurs as a subject; thus I never become merely an object of my own apprehension. The self always, and systematically, steps out of cognitive reach. Even if the reflecting self and the self reflected upon are numerically identical, I can never stand back and apprehend this identity since I shall always occur as a subject in my reflections, as well as an object. Qua subject I can never become an intentional object to myself. Yet it is qua subject that I have my essence (1993: 48, emphasis mine).

See also Zahavi (1999: 17).

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
speaking, absent dualism, they seem to be on par. So ontology can’t be the reason. Moreover, we know that the self is involved in every conscious act. That was the lesson of Descartes’ cogito.  

So ELUSIVENESS presents us with a bit of a puzzle: why is it that something that is known to be there during any conscious episode is something that, unlike the Ming vase, never makes an appearance on the phenomenological scene in any robust way (cf. Howell 2010)?

That ELUSIVENESS strikes us as strange only puts more pressure on the need to articulate its motivation. But in approaching this question, we would do well to recall that our charter does not include arguing for ELUSIVENESS. Instead, we are assuming ELUSIVENESS; our goal is to design a subjectivist theory that delivers self-consciousness in spite of ELUSIVENESS. Nonetheless, the sheer breadth of theorists who have endorsed something like this thesis (fn. 88) is reason enough to canvass their motivations for doing so, even if only briefly. I do not suggest that all of these theorists would endorse ELUSIVENESS as defined—seeking a plausible thesis that will satisfy everyone will end with us satisfying no one—but I believe that taking a look at what these theorists have to say will be worthwhile all the same.

We start with Hume, arguably the most famous proponent of ELUSIVENESS. For Hume, ELUSIVENESS is simply phenomenologically evident. He makes an introspective observation, and assumes that his observation isn’t unique. Hume did profess to have no idea of the self, and so was puzzled over what it was we are supposed to be looking for when we look for a self.  

---

97 Pace Lichtenberg. See Williams (1978) or more recently Peacocke (2014).

98 Here’s the relevant text: “[F]rom what impression cou’d this idea be deriv’d? This question ‘tis impossible to answer without a manifest contradiction and absurdity; and yet ‘tis a question which must necessarily be answer’d, if we would have the idea of a self pass for clear and intelligible” (1978: 251).

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
given his staunch empiricism, this point cannot be readily separated from his phenomenological one. So for Hume the case for ELUSIVENESS might seem to begin and end with phenomenology.  

Nonetheless, it is surely still reasonable to wonder why the phenomenology is the way it is. Howell (2010) has recently attempted to provide an answer. Howell points out that we are typically aware of objects by being aware of their intrinsic properties. Conscious states are properties of subjects. Howell argues that conscious states are not object-presenting, however. Even assuming such states are intrinsic, though had by subjects, the transparency of experience (§ 3.2) has the phenomenological ramification that conscious state are not presented as properties of subjects (2010: 474). Though Howell’s line is partially phenomenological—the appeal to transparency is an appeal to the way things seem—he is, at least on the face of it, doing more than Hume. Howell is appealing to a phenomenological datum (transparency) to explain why the self is elusive; the self is elusive because the very properties awareness of which would make the self not elusive are not object-presenting by their nature (ibid: 476). According to

99 Again, I am not claiming that Hume ascribed to ELUSIVENESS, only that he ascribed to a version of the elusiveness thesis. That said, there is at least some textual evidence that Hume did ascribe to our thesis (e.g. Butchvarov 1979: 251). I won’t wade into any historical exegetical waters, however.

100 Why intrinsic properties? Here is what Howell says:

The property of being a recording of Brahms’ second piano concerto is an extrinsic property of my record: it obviously has that property only given the existence of Brahms, as well as the existence of record players, lathes, etc. But this extrinsicality has epistemological and phenomenological implications as well. The property only evidences itself when the record is on a turntable that is amplified and generating sound waves through speakers. When one listens to the concerto, the record itself seems to disappear, despite the fact that one is in some sense perceiving one of its properties. The roundness of the record, in contrast, clings to the record itself. The record cannot ‘disappear’ when one perceives its roundness (2010: 474, emphasis mine).

Howell’s point is that mental states are an exception to the rule: though intrinsic, they are not object-presenting because that would upend their function, and hence ELUSIVENESS is true.
Howell, if conscious states were object-presenting, they would present ourselves instead of objects in the world, and thus fail in their purpose (ibid).

Not all defenders of ELUSIVENESS rely on its supposed phenomenological laurels, however. For others it is a conceptual truth. Schopenhauer remarked that the rejection of ELUSIVENESS is a “monstrous contradiction”: “[t]hat the subject should become an object for itself is the most monstrous contradiction ever thought of.” 101 Similarly, D.W. Smith claims that in being an object of experience, one is ipso facto not given as a subject of experience (1986: 152).

Gilbert Ryle is another defender of ELUSIVENESS, or at least something very similar. He says:

Even if the person is...momentarily concentrating on the Problem of the Self, he has failed and knows that he has failed to catch more than the flying coattails of that which he was pursuing. His quarry was the hunter (1949: 198).

Ryle was a behaviorist. It is unlikely, then, that he would have defended ELUSIVENESS only on phenomenological grounds, and he does not seem to think the issue is conceptual. Instead, his defense rests on the notion of “higher-order actions.” The idea is this. Higher-order actions are actions that are concerned with, or are operations upon, other actions. Self-consciousness is a species of higher-order action. And that’s precisely why the “I” is elusive. If we try to be aware of ourselves now, we perform a higher-order action now. However, Ryle argues, what we will attempt to have done (be aware of ourselves as subjects) cannot be done. For at every attempt, an extra fact is systematically added; every attempt at self-consciousness adds yet another act of self-consciousness. For Ryle, trying to have an experience of oneself as the very thing that does

---

101 Cassam (1997: 2) cites Christopher Janaway as the source for this attribution to Schopenhauer.
the experiencing at the present moment of experience is—to borrow a phrase from Kriegel (2003a: 479)—like “trying to hop on one’s shadow.” You are always just out of reach.102

Finally, Sydney Shoemaker (1968) gives an altogether different sort of argument. Call it the identification-free argument. According to Shoemaker, object-consciousness is identification-dependent: to identify a particular object x as being y, we must be aware of x as having a number of identifying properties which are true of y. For instance, to identify a bowler hat x as my bowler hat y, I must (i) know that it is true that ‘y is F and F*’ (e.g. that it is true that ‘y is black and has a scuffed rim’), and (ii) be aware of x as having F and F*. However, self-consciousness is fundamentally identification-free. Shoemaker argues by reductio. Consider the thought with the content ‘I am F’. If every instance of self-consciousness required self-identification—if self-consciousness is identification-dependent—then to know that I am F would require my picking out an object x that is F, knowing that I am the unique possessor of F, and then that I am x. But if knowing that I am the unique possessor of F also requires self-identification, we are off to a regress. To avert regress, self-consciousness must be identification-free, and hence not an instance of object-consciousness. If self-consciousness is not object-consciousness, ELUSIVENESS is true.103, 104

102 See Ramsey (1955) for critical discussion of Ryle on ELUSIVENESS.
103 Some proponents of subjectivism explicitly deny that self-conscious is identification-free. For example, Rosenthal—a proponent of unconscious subjectivism—claims, pace Shoemaker, that there is “thin” self-identification. According to Rosenthal’s HOT brand of HO theory, the content of each HOT refers to the subject in question (i.e. the subject in the mental state), but it does not explicitly describe the individual at the thinker of that HOT. If it did, a regress would ensue. Instead, Rosenthal cashes out thin self-identification in terms of the subject possessing a disposition to identify the individual that the HOT refers to as the subject who is thinking that HOT (2012b: 39).
104 Technically, this argument falls short of ELUSIVENESS. It does not establish that no instance of self-consciousness is an instance of object-consciousness; the argument only establishes that basic self-consciousness is not a form of object-consciousness. Still, it is quite clear that Shoemaker ascribed to ELUSIVENESS or something very much like it, since he thinks the best way to react to this reductio is to deny

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
4.2.3 THE OBJECT CONSTRAINT

Such is all I will say by way of motivating ELUSIVENESS. The next question is this: assuming ELUSIVENESS is true, what sort of constraints does it put on subjectivism? To some ears this might sound wrong-headed. ELUSIVENESS is not a constraint on subjectivism. What ELUSIVENESS shows is that subjectivism is false. Jesse Prinz (2012) seems to take this line. Prinz calls himself an eliminativist about the “phenomenal self” because, on his view, “among the various phenomenal qualities that make up an experience, there is none that can be characterized as an experience of the self or subject in addition to the qualities found in the perceived features of the world, sensations and emotions” (ibid: 214). But note that to speak of phenomenal qualities in this way is, on the face of it, to presume that if the self is phenomenally manifest, it must be phenomenally manifest as an object. So, strictly speaking, eliminativism does not follow. For it to follow, we need to establish the following conditional: if there is self-consciousness, then self-consciousness is a species of object-consciousness. Prinz provides no reason to affirm this claim.

---

that any form of self-consciousness requires identification, and hence is a form of object-consciousness (1986: 110). In any event, this point won’t matter for the subjectivist, for the subjectivist will insist that it is basic self-consciousness that is a constitutive feature of phenomenal consciousness.

Zahavi & Kriegel (2015) make a similar point against Prinz, albeit with respect to the so-called ‘for-me-ness’ of experience. Though Prinz does not explicitly mention for-me-ness as the target of his criticisms, Zahavi & Kriegel argue that for-me-ness is in any event immune; for Prinz’s criticism to be suitably extendable, the experiential reality of for-me-ness would have to amount to something like a “detachable self quale” (ibid: 39). Although I confess that I do not have a clear grip on what the phenomenal sense of for-me-ness is supposed to amount to, it seems clear that Zahavi & Kriegel have in mind something along the lines of the objection being voiced here.

Of course, there is a grammatical constraint that makes the description of self-consciousness a transitive form of consciousness—we are conscious of ourselves—and hence ostensibly a form of object-consciousness, that might seem to vindicate Prinz. Subjectivists have not been ignorant of this apparent grammatical tension. Zahavi (2006b: 23) points out that while grammatical strictures compelled Sartre to write we are aware of ourselves qua subject (“conscience de soi”)—thus suggesting ‘transitive self-consciousness’—he admitted that the use of ‘of’ (or ‘de’) is misleading for it suggests that the manner in which we are aware of ourselves qua subjects is structurally similar to the manner in which we are aware of objects. So while the ‘of’ may be a grammatical requirement, that it is does not carry any further import.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
Could one supply an argument for Prinz’s eliminativism? I’m doubtful. As noted, common
to cases of object-consciousness is that what is manifest as an object (a Ming vase, garlic) are
available for demonstrative reference, reference that doesn’t proceed by an object’s satisfying
or ‘fitting’ some descriptive condition). Now the view on hand is not that an instance of object-
consciousness has the phenomenology it has because the object is available for demonstrative
reference. I think it’s the converse: it is because object-consciousness has the phenomenology it
has that the object is available for demonstrative reference.\(^{107}\) But still, the relationship between
consciousness and demonstrative reference can be tricky. For example, John Campbell (2002) is
well known for defending the following thesis:

\[(1) \text{ If } O \text{ is available for demonstrative reference by a subject } S, \text{ then } S \text{ must be able to}
\text{consciously attend to } O.\]

And if, as seems plausible, being able to be attend to \( O \) requires the ability to be conscious of \( O \),
Campbell also holds:

\[(1)^* \text{ If } O \text{ is available for demonstrative reference by a subject } S, \text{ then } S \text{ must be able to}
\text{conscious of } O.\]

ER is threatened neither by (1) nor (1)*. But the following thesis does threaten ER:

\[(2) \text{ If a subject } S \text{ is conscious of } O, \text{ then } S \text{ must be able to demonstratively refer to } S \text{ where}
O = S \text{ (and where ‘conscious of } S \text{’ is read as ‘conscious of } S \text{ qua subject’).}\]

Does anyone ascribe to (2)? I’m not sure, at least not explicitly. Tye (2010), however, does ascribe
\(^{107}\) This is similar to Smithies’ (2011) view, although he does not explicitly discuss object-consciousness.

on its own. This is likely why Sartre placed the ‘of’ inside parentheses, speaking of “conscience (de) soi”
(Sartre 1943: 22, 1948: 62; also cited in Zahavi, ibid). The natural move here—the move to an ‘adverbial’
translation—will be discussed in below in § 4.2. My own view (§ 4.4 - 4.5) is that we should take this
grammatical constraint seriously, but not in a way that vindicates Prinz. In short—and pace Zahavi—I don’t
think it follows from the fact that self-consciousness is a matter of being conscious of ourselves that self-
consciousness is a species of object-consciousness.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
to the following:

(2)* If a subject S is *visually* phenomenally conscious of O (i.e. if S *phenomenally sees* O), then S must be able to demonstratively refer to S.

Tye motivates (2)* by way of cases like the moth perfectly camouflaged on the tree. The moth is in plain view, and there is certainly some sense in which I see it. But I am not visually phenomenally conscious of the moth because I cannot even ask myself ‘what's that?’ on the basis of my experience (*ibid*: 413).

ER, of course, is no more threatened by (2)* then it is by (1) or (1)*; it’s not as if self-consciousness in the sense under consideration is matter of literally seeing oneself qua subject. And if Tye’s argument for (2)* extends beyond the visual cases at all, it likely only extends to other perceptual cases, like olfactory or auditory experiences. But that’s hardly enough to get us (2)—we don’t smell or heard ourselves qua subject, either. Moreover, if all that’s required of us here is a mechanism for singular reference, demonstratives, of course, are not the only one; indexicals, like ‘I’, do just as well.

Thus it is not right to say that ELUSIVENESS establishes that subjectivism is false. What it does is place a *constraint* on subjectivism. The constraint tells us what the subjectivist needs to say if she *accepts* ELUSIVENESS. We can frame the constraint as follows:

**THE OBJECT CONSTRAINT:** If the self is phenomenally manifest in experience, it is not manifest as an object of experience.\(^{108}\)

*THE OBJECT CONSTRAINT* rules out any theory on which all consciousness is object-consciousness. To be clear, this is not a claim about *content*. Proponents of ELUSIVENESS sometimes flirt with this

---

\(^{108}\) The closest I have seen to an explicit mention of this constraint can be found in Legrand (2007: 588).

**SELF-CONSCIOUSNESS DESPITE ELUSIVENESS**
suggestion (e.g. Smith 1986: 152; Drummond 2006: 208), but they are wrong.\footnote{For example, Drummond says: “[t]he language of ‘is represented’ suggests a view of intentionality that is more appropriate...for object-awareness and unsuited for self-awareness” (2006: 208).} The expression ‘as an object’, again, functions as a \textit{phenomenological description}. It points to the sense in which those things that are available for demonstrative reference—what we are calling ‘objects’—are phenomenologically manifest in a specific and particularly salient way. For \textsc{the object constraint} to have any bearing on whether the self is represented in experience, we would need some argument for the claim that anything that is not available for demonstrative reference is thereby \textit{not} represented in experience. I have not seen such an argument in the literature, and I have no idea what such an argument would look like.\footnote{So the expression ‘as an object’ can also be distinguished from Shoemaker’s (1968) usage, which concerns the epistemic basis of certain first-person judgments. While I don’t claim that the account to be presented lacks connections with Shoemaker’s classic discussion of immunity to error through misidentification, I don’t explore them here.}

That said, the severity of \textsc{the object constraint} is hard to gauge. While virtually all theories of consciousness take paradigmatic instances of object-consciousness (e.g. visual experiences of Ming vases, olfactory experiences of garlic) as their central explananda, it is yet to be seen whether we need to reinvent the wheel, or whether an extant theory can be extended to accommodate self-consciousness despite \textit{elusiveness}. (For the record, I’ll take the latter approach, extending standard representationalism as the basis for ER).

Note, however, that it is not unreasonable to think that we already recognize, at least pre-theoretically, forms of consciousness that are not object-consciousness. Moods like ‘free-floating’ anxiety or elation can seem to be entirely undirected, lacking objects altogether. The suggestion here is not that the subjectivist who grants \textit{elusiveness} ought to model self-
consciousness by way of our theorizing about free-floating moods. I won’t. The point about moods is that it shows that we already have a grip on what a non-objectual form of consciousness might look like. And that’s a good thing.  

In any event, I’ll grant that The Object Constraint at least prima facie portends only one way out for the subjectivist, viz. an adverbial treatment of self-consciousness. Here self-consciousness becomes a matter of how we experience, not what we experience. In the next section, we’ll examine and ultimately reject this option. Doing so will help us get a better grip on what it would take to truly meet The Object Constraint, before turning to ER.

4.3  Adverbialism about Self-Consciousness

Zahavi (2006b; 2014) and Mark Rowlands (2013) advance adverbial theories of self-consciousness. Traditional adverbialism (e.g. Ducasse 1942; Chisholm 1957) was employed in the service of rejecting an Act-Object account of perceptual experience, specifically to circumvent the need to posit mental concreta (e.g. sense data) to which perceptual experiences are related. The Act-Object account is comprised of two claims. The first is called the relational thesis: perceptual experience consists in a subject standing in the relation of sensing to some object. The second is called the phenomenal thesis: the (represented) qualities of the object of experience constitute the phenomenal character of perceptual experience.

The adverbialist rejects both the relational thesis and the phenomenal thesis. Suppose Carmichael has a perceptual experience of red. Although the surface grammar of ‘Carmichael perceives red’ casts perception as a relation between two concrete particulars (Carmichael and

---

111 Though for a view on which this is not the case for moods, see Mendelovici (2014).
a red thing), we can paraphrase these perceptual ascriptions in a way that suggests a non-relational metaphysic. For instance, it can be paraphrased into ‘Carmichael perceives redly,’ which suggests a state of affairs in which a particular (Carmichael) instantiates a property (perceiving) and that property instantiating some other property (redly), not a relation between Carmichael and redness. (Compare ‘Carmichael walks briskly’). If we take this paraphrase seriously, the relational thesis is false. The falsity of the phenomenal thesis does not follow from this in a strict sense, but it is sensible. If perceptual experience does not consist in our standing in a relation to some object, then the phenomenal character of that experience is plausibly not the (represented) qualities of the object of that experience.

The motivation for adverbialism about self-consciousness comes from THE OBJECT CONSTRAINT.112 By rejecting the Act-Object Account with respect to self-consciousness, it might seem like we do not need to treat the self as an object of consciousness at all. Instead of ‘Carmichael is self-conscious of his perceiving the apple,’ we can paraphrase and adopt: ‘Carmichael is self-consciously perceiving the apple.’ The former reports a state of transitive self-consciousness; it reports the occurrence of a second-order state whose object is Carmichael’s perceiving the apple. The latter reports a state of intransitive self-consciousness; it reports the

112 Technically, Rowlands does not discuss ELUSIVENESS, but in describing Sartrean “non-positional consciousness,” he sees an extension to “pre-reflective” self-consciousness. For example, he says “If we tried to explain [self-consciousness] on the act-object model supplied by reflective consciousness, we might look for an introspectively discernible feature of my experiences—the property of mine-ness or being mine—in virtue of which they present themselves to me in this way...However, the same phenomenon can also be explained in adverbial terms...And this adverbial modification would be what the non-positional awareness of experiences as mine consists in” (ibid: 535). See also Amie Thomasson (2000) for a similar discussion with extensions to self-consciousness.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
occurrence of first-order state that occurs self-consciously. If self-consciousness is a matter of how we experience or a way of experiencing rather than what we experience, it seems we can indeed be self-conscious despite ELUSIVENESS.

Is adverbialism viable? I don’t like the adverbial approach in general for several (familiar) reasons, and there is little reason to that these reasons don’t carry over to adverbialism about self-consciousness. For present purposes, however, we’ll restrict ourselves to assessing adverbialism’s merits only with respect to the problem at hand, viz. delivering self-consciousness despite ELUSIVENESS.

The problem starts with a suspicion. The suspicion is that the adverbial approach amounts to a mere grammatical trick; although we can say things like ‘Carmichael is self-consciously perceiving the apple,’ it is not clear what this means. There is—again, at least a suspicion—that it is cheating somehow. For instance, consider what Dan Zahavi tells us:

> [O]bject-consciousness necessarily entails an epistemic divide between that which appears and to whom it appears, between the object and subject of experience, and that is why object-consciousness (and the transitivity principle) might be singularly unsuited as a model for self-consciousness proper. (2014: 16)

Leaving aside the claim that there is an epistemic ramification to object-consciousness that explains why ELUSIVENESS is true, Zahavi is just expressing THE OBJECT CONSTRAINT. But then on a positive note, he later adds:

---

113 See Kriegel (2003c, 2009) for more on transitive and intransitive self-consciousness. However, Kriegel ultimately treats intransitive self-consciousness as a form of “peripheral inner” object-consciousness (2009: 102), and thus does not meet THE OBJECT CONSTRAINT (see Zahavi 2014: 16 for discussion).

114 Traditional adverbialism faces some well-known trenchant challenges (e.g. Jackson 1977), which may or may not carry over here. Also, while I am neutral here on the prospects for naturalizing consciousness, if one thinks that reductions must be explanatory, there might be reason to think that adverbialism makes naturalizing consciousness harder. This is because treating consciousness as an intrinsic property makes functional reduction—a fairly standard form of explanatory reduction—a challenge, if not an outright impossibility (Kriegel 2011). As such, this will make adverbialism unattractive to anyone with such hopes.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
The self can be given, can be real and possess experiential reality even if it does not appear as an object of, or an item in, experience. The experiential self...doesn’t denote an experiential item or object, but the very first-personal mode of givenness...It refers to the first-personal presence of all my experience content; it refers to the experiential perspectivalness of phenomenal consciousness. When I have experiences, I, so to speak, have them *minely*, [and having them minely] is an adverbial modification of the act [of experience] rather than the object (*ibid*: 22, emphasis mine).

I am not sure what Zahavi means by the “first-personal mode of givenness” of experience, beyond the fact that experience is subjective. Indeed—although his usage is inconsistent—Zahavi occasionally employs “subjectivity” as a substitute for “first-personal mode of givenness” (2006b: 126; 2014: 20). If that is what Zahavi means, however, it seems like we have a category mistake. For the subjectivity of experience is not something that can, on its own, make any phenomenal difference. Subjectivity is a formal feature of experience; to say that an experience is subjective is just to say, trivially, that the experience is had from a point of view that is only occupiable by the very thing that is having that (token) experience. That an experience is subjective may have epistemic implications—I’ll leave that aside here—but even if it does, subjectivity itself is not something with which we can be acquainted.

But suppose Zahavi does *not* mean to invoke subjectivity in a mere formal sense. The adverbial approach still has a serious problem. The problem is that it is unclear why we should describe Carmichael’s property of *self-consciously perceiving the apple* as the (or a) property of *being self-conscious*. To see this, consider the question of individuating experiences unique to the traditional adverbialist. We know that reddish visual experiences are different from greenish visual experiences. But what makes it such that my visual experience is a *reddish* visual experience as opposed to a *greenish* (or some other) experience? The adverbialist claims that these are different manners of sensing: the reddish way, and the greenish. But this doesn’t do...
much aside from recast the question: what makes one manner of sensing the red manner, as opposed to the green manner? The proponent of an Act-Object Theory, by contrast, has a straightforward answer. She will just point to features of the object of experience; red experiences are reddish because they are of things that are (represented as) red, and green experiences are greenish because they are of things that are (represented as) green.

The adverbialist clearly cannot go this route and remain an adverbialist. So what can she say? Perhaps the most obvious option—suggested by Tye in his adverbialist days—begins with correlations between manners of sensing and the physical objects that cause those sensations. Following Sellars, for Tye, “manners of sensing are conceived as resembling and differing from one another in ways systematically analogous to the ways in which their ‘corresponding’ physical object, i.e. their normal causes, resemble and differ from one another” (1975: 139). What makes a reddish experience reddish instead of greenish, and what explains the relationship between reddish and greenish experiences, then, is parasitic on the relationship that holds between red and green physical objects. That is, those things that typically cause experiences, but not (of course) to those things to which we stand in any experiential relation.

I do not see how the subjectivist can avail herself of Tye’s solution. On subjectivism, self-consciousness is a constitutive feature of phenomenal consciousness. So, on subjectivism, every instance of phenomenal consciousness will be an instance of self-consciousness. But then the property self-consciously perceiving will correlate with just about everything that is capable of causing an experience at all. Every time I have a reddish perceptual experience, and every time I have a perceptual experience at all, I am—by the lights of the adverbial subjectivist—self-consciously perceiving. The worry suggested by this is that the term ‘self’ in ‘self-consciously
perceiving’ is just a tag; we are supposed to be articulating a theory concerned with self-consciousness, but for all the adverbialist has told us, we have no reason why the property self-consciously perceiving should track anything resembling self-consciousness at all.

Can the adverbial subjectivist say anything to speak to this worry? I’m doubtful. But I do not want to stake my success on this claim. Instead, I want to suggest that these aforementioned considerations point to a constraint on THE OBJECT CONSTRAINT itself: if possible, we should treat self-consciousness as the surface grammar suggests, i.e. as consciousness of a self. And since only selves are conscious of things, self-consciousness has to consist in consciousness of a self by a self. So, on the face of it, self-consciousness consists in the obtaining of a relation—a reflexive relation. It is a relation we bear to ourselves. This is its structure. Yet we lose this when we invoke an intransitive form of self-consciousness and take it seriously in metaphysical terms. True, the traditional adverbialist was motivated by the well-founded desire to do away with non-physical sense data. But in doing so, she also ridded us of something that is unproblematic, viz. the relational structure of experience. This is why intentionalism ultimately carried the day. Intentionalism rid us of sense data while retaining the apparent structure of experience. It gave us the best of both worlds.  

In short, the adverbialist ‘solution’ to the problem of delivering self-consciousness despite ELUSIVENESS is an overreaction. The trick to delivering self-consciousness despite ELUSIVENESS is to

---

115 It is sometimes said that the Act-Object Conception is exclusive to sense datum and naïve realist theories. I think this is false. An intentionalist (or representationalist) can accept the Act-Object account. For one thing, the intentionalist might say that the content of experience is object-dependent. But more generally, the intentionalist will readily grant that experience consists in our standing in a relation to some proposition or properties (we are aware of objects when those properties are instantiated). Given our wide reading of the word ‘object,’ I think it is appropriate to say that the intentionalist ascribes to the Act-Object Account, even if in a weaker sense than the naïve realist.
figure out how to maintain its apparent reflexive structure while obeying the **Object Constraint**—
that is, by finding some principled way in which the self can be phenomenally manifest without
being manifest as an object. This is what makes the project so challenging. Adverbialism only
*adds* to the challenge. Think of it this way. Suppose we stipulate that there is some property \( F \).
And suppose we add that whenever we instantiate \( F \) we are ‘self-conscious’ despite **Elusiveness**.
We will have done nothing to advance things for the subjectivist if we have no principled reason
to claim that when we instantiate \( F \) we are truly self-conscious.

Zahavi et al. will no doubt reply that ‘(strong) self-consciousness’ in not a monolithic
concept, and that the richer the conception of self-consciousness, the less plausible it is that self-
consciousness is a constitutive feature of phenomenal consciousness. Little can be said for
delivering on the promise of self-consciousness despite **Elusiveness** yet ending up with a version
of subjectivism that is wildly implausible.\(^{116}\)

I don’t deny that self-consciousness is a many-layered phenomenon. Nor for that matter do
I deny (or at least I don’t assert the negation of) the claim that at least *some* conceptual resources
are required for all forms of self-consciousness. The problem, rather, is that if the conceptual
resources are *too* high, there will be creatures that we pre-theoretically take to be phenomenally
conscious that are in fact not because they *lack* such conceptual resources.\(^{117}\) So the conclusion

---

\(^{116}\) What are the wildly implausible forms of subjectivism on these grounds? This is hard to definitively
answer in the abstract and from the armchair, since conceptual sophistication is largely an empirical
matter, but it plausible that any form of subjectivism on which the relevant form of self-consciousness
was grounded in the possession of a theory of mind (e.g. Baron-Cohen 1989; Gopnik 1993), or in a capacity
to think of various experiences as belonging to one and the same subject (e.g. Cassam 1997), or in our
capacity to develop self-narratives (e.g. Dennett 1991), will certainly run into trouble.

\(^{117}\) Carruthers (2000) is a subjectivist who embraces the claim that most non-human animals are *not*
phenomenally conscious because this requires a ‘theory of mind’ consumer-system to transform first-
order perceptual contents into higher-order ones.
Zahavi draws is correct in spirit; if subjectivism is to be viable at all, we need something “logically and ontogenetically more primitive”—a “minimal” or “core” notion of self-consciousness—than the forms of self-consciousness just adduced (ibid: 14). But this point does not change the issue: “minimal” or not, F still needs to be deserving of the name.\textsuperscript{118}

By contrast, a form of subjectivism that maintains the apparent reflexive structure of self-consciousness does not have this problem: F is the property we instantiate when we stand in a suitable relation to ourselves. This, I grant, makes addressing ELUSIVENESS harder. But it doesn’t make addressing ELUSIVENESS a non-starter (pace Zahavi 1999, 2006a, 2014; Musholt 2015). That, in any event, is what I’ll try to show. In the next section, I’ll begin by sketching the basic idea at the foundation of ER. Having done that, I’ll fill in the details in § 4.4.2.

4.4 THE VIEW

4.4.1 LONERGAN ON SELF-CONSCIOUSNESS

Bernard Lonergan is the inspiration for ER’s foundation. Here’s the inspirational bit:

Objects are present by being attended to, but subjects are present (to themselves) as subjects, not by being attended to, but by attending. As the parade of objects marches by, spectators do not have to slip into the parade to be present to themselves (1967: 226).

\textsuperscript{118} Nor does the fact that Zahavi is attempting to capture a Sartrean-inspired pre-reflective self-consciousness. Zahavi’s mistake is to treat pre-reflective self-consciousness as a form of non-transitive self-consciousness. For Sartre, pre-reflective self-consciousness amounted to ‘non-thetic’ or ‘non-positional’ self-consciousness. But there is nothing about the fact that the self is not phenomenally manifest in a positional manner or as an object which requires us to say that when the self is phenomenally manifest, it’s being so is a non-relational affair. For more on Sartre’s views about self-consciousness and their relation to some of the distinctions drawn here, see Longuenesse (2008).

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
Lonergan is making two claims. The first, in effect, is an endorsement of THE OBJECT CONSTRAINT. I take it that this is (or is near enough) what he means when he says “spectators do not have to slip into the parade to be present to themselves.”

The second claim is a (rather skeletal) thesis concerning how we are self-conscious despite ELUSIVENESS. Doing without Lonergan’s restriction to attention, the basic idea (we’ll suppose) is that subjects are present (to themselves) as subjects, not by being given as objects of experience themselves, but by being object-conscious, presumably in a special sort of way.

I cannot rule out that Lonergan had in mind an intransitive use of consciousness where the right sort of ‘way’ qualifies the act of awareness, or something else that might suggest a lapse into adverbialism. He doesn’t tell us. But what Lonergan really meant is not the issue. I am not interested in Lonergan exegesis. Instead, I am interested in taking the sketch of a solution that Lonergan does provide, and developing it into something concrete, and something that presents an advance over the adverbial approaches we have thus far considered.

---

119 Similar sentiments have been expressed elsewhere by Evans (1970), Deikman (1996), and Sass (1998), and even William James (1890). For instance, Deikman claims “we know the internal observer not by observing it but by being it” (1996: 355). This sounds a bit like what Lonergan is saying, but again ultimately it is too hard to be sure, since the details of Lonergan’s proposal are sparse. (Lonergan’s remarks about the possibility of self-consciousness despite ELUSIVENESS come in passing, in the context of a larger discussion of his ‘cognitional theory’ and different types of human knowledge).

120 Another nice feature of Lonergan’s thesis is that it comports with a claim made by Hume often missed by espousers of ELUSIVENESS. Hume does not only tell us that he “never can observe any thing but the perception”. He also asserts, just prior, the seemingly incompatible claim that he can “never catch myself at any time without a perception”, i.e. a perception “of heat or cold, light or shade, love or hatred, pain or pleasure” (1739/1978 I.vi.6). Here Hume does not deny the possibility of self-consciousness outright, but only insists that self-consciousness be accompanied by concurrent perception and sensations. Lonergan cannot absolve Hume from contradiction, but he might be able to vindicate this latter claim. For by Lonergan light’s, Hume was right: we are self-conscious by being suitably object-conscious, and thus a fortiori cannot be self-consciousness without being object-conscious, or as Hume would put it, a perception heat or cold, light or shade, and so on.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
So, taking Longeran in spirit, I suggest that self-consciousness, understood as the phenomenal manifestation of the self (as a subject of experience) is somehow embodied in object-consciousness. Yet despite being so embodied, it is not itself an instance of object-consciousness—the self is not manifest as an object of experience.

4.4.2 **Egocentric Representationalism**

ER cuts across two main theoretic classificatory divides. First, ER is an egocentric theory. To describe ER as an egocentric theory is just say that it is concerned with strong rather than weak self-consciousness. Second, ER is a species of representationalism. There are many forms of representationalism. It is fairly standard to restrict ‘representationalism’ as referring to FO theories. But this restriction is unnecessary. On our usage, to say that ER is a representational theory is just to say that it identifies phenomenal character—that property that all and only conscious states have—with a certain type of representational content. There is no stipulation that this content is a first-order content (as on FO representational theories), or a higher-order content (as on HO representational theories). ER can be cast as a FO theory or as a HO theory. This is a virtue. If you like subjectivism, but are inclined towards a FO framework, you can adopt ER. Alternatively, if you are inclined instead towards a HO framework, you can also adopt ER.

ER is also neutral between *pure* or *impure* representational theory, i.e. whether it must appeal to manners of representation, like functional characterizations of the representing state.\(^\text{121}\) Relatedly, ER does not say anything about what makes a mental state conscious. ER only

---

\(^{121}\) Most FO representational theories are impure. To the best of my knowledge, the closest we have to a pure representational theory is David Bourget’s (2010) PURE theory, according to which phenomenal states are simply *underived* intentional states, i.e. those states that are intrinsically representational. According to Tye’s (2000) PANIC theory, phenomenal states are intentional states that are poised,
has something to say about the content of conscious states. This is where ER’s two components meet. Its representational component: phenomenal character is exhausted by representational content. And its egocentric component: the content is essentially \textit{de se}. What follows is an attempt to articulate the structure of those contents.

Let’s begin by returning to Lonergan. He tells us “subjects are present (to themselves) as subjects, not by being attended to, but by attending” to the “parade of objects.” To build this into something approximating a real theory, it will be helpful to distinguish two questions. First:

\textbf{How possibly?} How can the phenomenal manifestation of the self as a subject of experience come about by our being object-conscious (in “attending...to the parade of objects”), without the self being an object of experience in the process (without “slip[ping] into the parade to be present to themselves”)?  

To answer the \textbf{how possibly} question is to understand the idea, alluded to earlier, that self-consciousness is somehow embodied in object-consciousness. The second is:

\textbf{Why necessary?} Why is it that, necessarily, the self cannot be an object of consciousness? (Why is it that, necessarily, selves do not “slip into the parade to be present to themselves”?)

To answer the \textbf{why necessary} question is to explain why \textit{Elusiveness} is true. Doing so is not, strictly abstract, and non-conceptual. The appeal to the property of being poised makes Tye a \textit{functionalist} (\textit{ibid}: 62). The reason most are impure is that it is likely that anything that can be consciously represented can also be unconsciously represented; all kinds of contents can go on in the dark (Kriegel 2002; Chalmers 2003). That said, the implausibility of a pure form of FO representationalism is typically premised on the notion that we can unconsciously represent shape, size, and color properties, and so-called ‘environmental properties.’ But a pure representationalism may end up being viable if we expand the base of content properties beyond the common fair as ER does.

\footnote{This is only a possibility claim because I am building a model of self-consciousness despite \textit{Elusiveness}; I’m not claiming (or at least not primarily claiming) that the model accurately represents the world.}
speaking, incumbent on the subjectivist. Accepting ELUSIVENESS only forces the subjectivist to show how one might meet THE OBJECT CONSTRAINT.\(^{123}\) The how possibly question speaks to that challenge. Explaining why ELUSIVENESS is true, once we have accepted it, is a further issue. So I’ll address the how possibly question first, and only then move to consider the why necessary question.

Consider an ordinary instance of (visual) object-consciousness: the experience I have when I see a ripe tomato in good light. When I have such an experience I am aware of the tomato as being red, round and bulgy. Something similar will be true of any instance of object-consciousness. Any determinate instance of object-consciousness will involve the awareness (and so representation) of a whole host of qualitative properties, and it is our awareness of these properties (and objects, when the experience is veridical) that individuate distinct episodes of object-consciousness.

But this, according to ER, is not all we are aware of: we are also aware of the tomato as being other than ourselves. Moreover, this is true of all instances of object-consciousness. In other words, on ER, if \(x\) is an object of experience, then \(x\) is represented as standing in a certain relation to oneself, viz. the other than relation.\(^{124}\)

\(^{123}\) THE OBJECT CONSTRAINT, recall, said that if the self is a phenomenally manifest aspect of experience then it is not manifest as an object of experience. Note also that it is not incumbent on us to argue for the antecedent of this claim. We are, again, only trying to articulate the structure of phenomenological content on the assumption that the conjunction of subjectivism and ELUSIVENESS is true. We are not trying to show how an unconscious state becomes a conscious state with that phenomenology.

\(^{124}\) In general, representing relations is a common affair in perceptual experience, as when we represent spatial contents (e.g. to the front, or to the left of). Sometimes, these contents are understood in terms of self-involving relational properties—e.g. representing the tree as being to the left of me (e.g. Schwenkler 2014). It is unlikely, however, that an appeal to spatial contents can help the subjectivist who accepts ELUSIVENESS. For one thing, the talk of self-involving relational properties is misleading. Spatial contents are best understood as the representation of relation between locations, and reference to the self (as opposed to the body) arguably does not pick out a unique location. But even if it were not the case, the
It is worth noting that there is a sense in which this tallies well with our earlier remarks concerning the connection between object-consciousness and availability for demonstrative reference. To borrow an example from Susanna Siegel (2002: 4), suppose Carmichael looks at a chameleon that blends in perfectly with the area surrounding. It is a matter of controversy whether in such a case Carmichael sees the chameleon. But even if Carmichael does see the chameleon, Siegel points out that Carmichael’s seeing the chameleon does not suffice to make it available to Carmichael as an object of demonstrative reference, and thus an object of experience in our sense (ibid). However, if Carmichael can differentiate the chameleon from its surroundings, Carmichael can then refer to it demonstratively if he so chooses.

Suppose now we switch cases. Instead of a chameleon, Carmichael has a visual experience of an undifferentiated blue expanse, occupying the entirety of his visual field. (I’ll discuss cases like these—‘ganzfeld experiences’—again in § 4.5). Although atypical, the undifferentiated blue expanse is, I propose, an object of experience just like anything else. However, if that’s the case, we should be able to refer to it demonstratively. ER explains this nicely: even though there is no other object of experience from which one can differentiate the blue expanse, one can still differentiate it from oneself, viz. by representing it as other than oneself.

representation of self-involving relational properties would likely still not help the subjectivist. There are at least two reasons for this. First, by accepting ELUSIVENESS, the subjectivist is forced to take THE OBJECT CONSTRAINT seriously. That amounts to explaining how the self can be phenomenally manifest as a subject of experience without being an object of experience. But there is little reason to think that the representation of spatial relations like to the right of or in front of can help in this regard. Second, it is also controversial that perceptual experience necessarily contains self-related information, whether or not such self-related information is genuinely about the self (see, e.g. Perry 1993, Campbell 2002). But even if some perceptual experiences involved self-locating content, this won’t cut it; if such content is to help, all perceptual experiences must have such content.

Siegel (ibid) cites Dretske (1969) in this regard, who claims that visually differentiating an object O from its surroundings is generally necessary in order to see O.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
If this claim is true, an answer to the how possibly question comes swiftly. For now to be object-conscious is to be (at a minimum) conscious of something as other than oneself. But to be conscious of something as other than oneself is to be conscious of one’s self, i.e. to be self-conscious. It is to be conscious of one’s self in much the same way as being conscious of the tree as to the left of the fence is to be conscious of not just the tree, but the fence as well. So our earlier talk of ‘embodiment’ comes down not simply come down to the fact that self-consciousness is entailed by object-consciousness. Rather, self-consciousness is embodied in object-consciousness in the stronger sense that both forms of consciousness occupy two poles of a single represented relation, viz. the other-than relation. To be object-conscious is to already be self-conscious. Lonergan makes something like this very point when he says: “[subjects] are present to themselves by the same watching that, as it were, at its other pole makes the parade present to them” (1967: 210). The members of the “parade” are, in our terms, the objects of experience. The key, however, is that we are phenomenally manifest (to ourselves) as subjects “by the same watching” in virtue of which something is manifest as an object of experience. And that’s precisely what ER says.

Nor, in saying this, do we violate ELUSIVENESS. The other-than relation is irreflexive: one cannot represent something as other, and in doing so, be aware of that very thing as oneself, even if that thing is—as a matter of fact—oneself. Thus, if every object of consciousness has the property of being represented as other than oneself, then one cannot experience oneself as an object. And that’s a way of delivering self-consciousness despite ELUSIVENESS.

In sum, the view on the table is this. Let S be an arbitrary subject of experience, and let the expression ‘\(\mathcal{O}OBJECT\_x\)’ be read as ‘x is phenomenally manifest as an object,’ and the expression

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
'SUBJECTx' be read as ‘x is phenomenally manifest as a subject.’ To say that x is phenomenally manifest as an object is to say that x is an object of experience (i.e. we are object-conscious with respect to x). To say that x is phenomenally manifest as a subject of experience, is to say that x is self-conscious, and in a way where x is thereby not an object of experience. Returning to our stock example, suppose now that Carmichael has a visual experience e of a Ming vase called Vasey. ER is the conjunction of the following four theses. Here are the first three:

(1) If e is conscious, then $\phi$ SUBJECTCARMICHAEL of e to Carmichael.
(2) $\phi$ SUBJECTCARMICHAEL of e if and only if $\phi$ OBJECTVASEY of e to Carmichael.
(3) $\phi$ OBJECTVASEY of e to Carmichael only if e represents Vasey as other than Carmichael.

Thesis (1) just points out that ER is a species of strong conscious subjectivism. Thesis (2) draws out the claim that object-consciousness and self-consciousness are inherently linked. But unlike object-consciousness, representing something as other than oneself is all self-consciousness consists in, at least under ELUSIVENESS. We do not represent ourselves—not, at least qua subject— as having various qualitative features like being a certain size, shape, and color. Thus, under ELUSIVENESS, what it is like to be self-consciousness is very different from what it is like to be object-conscious. And this is exactly as it should be. If the self were experienced as having distinctive features as in the case of object-consciousness (e.g. as having a certain shape, like Carmichael’s Ming vase), ELUSIVENESS would be a hard sell. The phenomenological self is thus akin to a nearly featureless locus of consciousness—simply as that thing which is experienced as being other than the object of experience.

What about the why necessary question? Here matters are more delicate. To speak to the why necessary question, we need to know what it is to be an object of experience. We need to
know, in other words, what it is that the tomato and the Ming vase have in common that makes them objects of experience. Relatedly, we also need to know why, because of this, the self necessarily cannot be an object of experience, at least on the supposition that ELUSIVENESS itself a necessary truth. Relevant here is ER’s fourth thesis:

(4) \[ \Box \neg \left[ \text{Carmichael is the subject of } e \land \Box \text{OBJECT} \text{CARMICHAEL of } e \text{ to Carmichael} \right] \]

Again, thesis (4) is true according to ER because the other-than relation is irreflexive. I cannot represent something as other than me and be aware of that something as me. So if (part of) what it is for an \( x \) to be an object of an experience (to be object-conscious with respect to \( x \)) is for Carmichael to have an experience \( e \) that represents \( x \) as other than Carmichael, then we have an answer to the why necessary question. But what makes something an object of experience is a substantive matter, and this answer won’t—at least here—be defended. I’ll note, though, that the proposal is far from outlandish, especially when we recall Siegel’s (2002) point about differentiation: if differentiation is key for demonstrative reference, and demonstrative reference is key for object-consciousness, then representing the other-than relation seems like it too will be key for object-consciousness. Whether the self must always be amongst the relata of the other-than relation is a further question.

All the same, the subjectivist can, as noted, rest easy with just an answer to the how possibly question, or at least in the present context. Moreover, this answer—that self-consciousness consists in representing oneself as standing in an irreflexive relation—is consistent with self-consciousness actually being reflexive, contra adverbialism. In having a visual experience of the tomato, I am self-conscious in virtue of the tomato being represented as other-

---

126 Again, it can’t be because the self is not an object; the self (or so we can suppose) is an object too.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
than myself. The other-than relation is the single Lonerganian relation: it is by representing this single relation (the “same watching”) that I am object-conscious (with respect to the tomato) and also self-conscious. But to say that there is an irreflexive relation \( R \), such that I am self-conscious only when I represent myself as standing in \( R \) to something, is not to say that I do not bear a reflexive relation in virtue of being self-conscious, or (more critically) that I am self-conscious in virtue of bearing a reflexive relation. The former claim is a claim about the sort of relation I represent myself as standing in, which does not, in itself, have any bearing on the relations I actually stand in when I am self-conscious.

And that self-consciousness is characterized this way, speaks to point noted earlier (§ 4.1) that ER can capture the spirit of pre-reflective self-consciousness so often discussed in the phenomenological tradition. By following Lonergan in tying self-consciousness to object-consciousness in such an intimate way, ER treats self-consciousness as an essentially outward looking affair (or more precisely, as an essentially other-directed affair). Here, self-consciousness is not a matter of reflecting inward, or introspecting, but a matter of representing ourselves as standing in a single, inherently externally directed Lonerganian relation—viz., the other-than relation. It is in this sense (and only this sense) that ER might be thought as giving a theoretical elucidation of pre-reflective self-consciousness.

Such is the view. No doubt more can be said about ER—and eventually, more ought to be said. But, again, our present aims are modest: we are only trying to give an account of the structure of phenomenal contents, not an account of how a conscious state gets those contents, nor how a mental state becomes conscious. Those are good questions, and ones that I hope will be taken as future challenges for ER.
Still, I submit that, even as it stands, ER should be taken seriously as providing the start of a solution to the problem of delivering self-consciousness despite ELUSIVENESS. It has, at the very least, the following merits. First, the solution is non-mysterious, at least insofar as representing relations is non-mysterious. It does not treat self-consciousness as a *sui generis* phenomenon (e.g. Frank 2007; Zahavi 2014). Second, this solution falls out of an account of object-consciousness. We are self-conscious merely by being object-conscious, without the self being manifest as an object of experience. So ER meets THE OBJECT CONSTRAINT. We also saw that, *pace* adverbialism, ER preserves the essential reflexivity of self-consciousness. Self-consciousness does indeed consist in bearing a relation to one’s self, it’s just that we represent ourselves as standing in an irreflexive relation to something when we do so. Of course, relations only hold between objects (in the non-neutral sense). Yet ER does not deny that the self—the subject of experience—*is* an object. What ER resists instead is the idea that bearing a relation to something makes that something an ‘object' in the relevant phenomenological sense.\(^{127}\)

### 4.5 ELUSIVENESS AND TRANSPARENCY

I proposed earlier that we judge ER as a theory for subjectivists, and that we assess its merits in terms of its ability to deliver self-consciousness despite ELUSIVENESS. I just articulated how ER

---

\(^{127}\) Are there any other competitors to ER, i.e. non-adverbial subjectivisms that accept ELUSIVENESS? Sebastián’s (2012) *self-involving representationalism* (or ‘SIR’) is a recent version of subjectivism that does not appeal to an adverbial metaphysic. However, Sebastián’s stance on ELUSIVENESS is unclear. He says that SIR is compatible with what Prinz calls “weak reading of Hume’s thesis” (2012: 214). On the *weak* reading of Hume’s thesis, there are I-qualia, but these are “nothing above and beyond the qualities of perception and sensation” (*ibid*: 214). On the *strong* reading, there are no I-qualia *period*. (Prinz, as we saw earlier, agrees with the strong reading). The problem is that even if SIR is compatible with Prinz’s weak reading of Hume’s thesis, SIR is not compatible with ELUSIVENESS. Since the qualities of perception and sensation are plausibly given as objects of experiences in the requisite sense, to suppose that self-consciousness can be reduced to these qualities is to contravene ELUSIVENESS.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
achieves the latter. To close, I want to briefly consider the extent to which ER is plausible on independent grounds. Of course, ER’s plausibility rests to a significant extent on the plausibility of subjectivism, but the plausibility of subjectivism is not up for grabs here. I propose instead to look at one issue that arises out the details just canvassed that might lead one to think that ER is not plausible, and show how the would-be proponent of ER might begin addressing it. My goal is to encourage the subjectivist to gravitate towards ER; if it also encourages the anti-subjectivism to gravitate towards ER, that’s great too. Doing so will bring us full circle thematically: it turns out that Transparency (§ 3.2), of all things, can tell us something about self-consciousness and Elusiveness, and the possibility of the former given the latter.

The simplest objection to ER is that even if it meets the object constraint without an adverbial metaphysic, we have little independent reason to think that its core theoretical posit—that every experience involves the representation of something as other—is true.

The way to approach this issue requires a detour. The detour consists in considering the following question: to what extent is experience reticent with respect to the metaphysical status of its phenomenal features? That is, does experience involve the appreciation of something as

---

128 So I am limiting myself to concerns particular to ER. As such, I won’t look at objections to representationalism. Is it possible to build on Lonergan’s insight without appeal to a representational-theoretic explanation? Perhaps. But absent treating consciousness as brute or primitive relation (akin to some naïve realist’s construal of the acquaintance-relation), I have no idea how to understand consciousness but as a species of representation. Moreover, in being a representational theory, ER makes naturalizing consciousness easier. (The thinking being, by many, that we have a decent antecedent grip on what it would take to naturalize mental representation [cf. Kriegel 2009]). By contrast, Zahavi seem to think that because self-consciousness cannot be assimilated to a form of object-consciousness, the former must be a sui generis and unstructured form of consciousness (1999, 2006a; cf. Frank 2007). But this is unsatisfying on purely methodological grounds. As Kriegel points out, we do not adopt a primitivist account of something at the beginning of enquiry (2009: 102). There might be cases where there is a principled reason to rule out a reductive account from the get-go, but Zahavi gives us no reason to think that this is one of those cases. So, it is, to my mind, worth it to at least attempt a reductive proposal, and ER permits that.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
having features that constitute membership in an ontological category? The categories I have in mind might be grouped into two families: those picked out by the predicates (i) is *objective*, is *physical*, is *mind-independent* on the one hand, and those picked out by the predicates (ii) is *mental* and is *mind-dependent* on the other.

A common starting point in addressing these questions is the observation that experience is *transparent*. In § 3 we examined a restricted, negative, disjunctive transparency thesis (‘W-TRANSPARENCY’) that, I argued, conflicted with *TRANSITIVITY*. However, it was also noted that many transparency theses also include *positive* claims. Return to Harman’s case of Eloise (see § 3.2).

The claim is not merely about what Eloise does not seem to be aware of, but also what she is aware of, i.e. features of the tree and its various relational features. The positive upshot of most transparency theses is that what we *are* aware of, when we undergo these experiences, is exactly what we seem to be aware of: nothing mental or mind-dependent, but simply various features of our environment. The idea, then, is to take this positive datum as speaking against various mentalistic or ‘mental paint’ views that treat experience as essentially constituted by, or related to, mental or mind-dependent phenomena.

But this is misguided. It is at best questionable that we can infer anything about the nature of experience from the mere fact that the phenomenological deliverances of experience do not seem to be mental (like Eloise’s tree). As Robert Schroer has pointed out, it does not follow from this that the phenomenological deliverances seem *non*-mental; instead, it is possible that they do not seem mental *and* they do *not* seem *non*-mental (2007: 401). Perhaps how phenomenal features seem is “silent” with respect to their ontological classification (*ibid*).
Schroer argues that we already (at least implicitly) acknowledge this. Take projectivism. We encountered projectivism in § 3.2. Recall that projectivism is an error theory: although Eloise experiences colors (and other qualities) as properties of bodies in her ambient environment, they are actually intrinsic properties of her experience. Yet despite this, the character of Eloise’s experience is not simply consistent with the phenomenological predictions issued by projectivism: it is exactly how we would expect it to be on projectivism (Frey 2013: 72). Projectivism could be true and an idealized version of Eloise would be none the wiser. Experience may make claims about where the relevant phenomenal features are located, but it makes no claim about whether these phenomenal features are mind-dependent or independent in their nature (Schroer 2007: 430).

What lesson shall we draw here? The lesson is not that the transparency phenomenon has no positive classificatory implications. The lesson is that the significance of its positive implications is modest. Experience is (to use Schroer’s term) largely *reticent*—how phenomenal features seem is “silent” on whether they belong in (i) or (ii). However, this does not mean, *pace* Schroer, that experience is *fully* reticent. Experience may be silent on *metaphysical* classifications, but this does not mean that it is silent on classifications *tout court*. The sense in which experience is not fully reticent, I claim, may help the advocate of ER make her case.

Here I follow Christopher Frey (2013). In looking for an appropriately parsimonious transparency thesis—i.e. one that is divested from these metaphysical commitments—Frey has us consider two cases. Both are rare, but according to Frey, are still telling. One case is *ganzfeld experiences* (ibid: 74). Ganzfelds are visual fields that are homogenous, entirely permeated with a single color. Frey uses the more familiar term ‘ganzfield’, but I am going to use ‘smanzfeld’
instead. Suppose now that we have a smanzfeld permeated with a determinate shade of blue. According to Frey, all we can say in such a case is that the blueness is appreciated as being “before” or as qualifying something “other” than one. We need not even experience the ganzfeld as distally located. So while there is no ontological classification (such as being mental or non-mental), there is still non-ontological classification, viz. as being other than oneself.

The second case Frey has us consider is spatially punctiliar experiences (ibid). When the occipital cortex is damaged, the visual field sometimes contracts. As Frey points out, what results in such cases is not mere darkness or ‘brain grey’ (as when external optical stimulation is lost). The visual system fails to register anything in the lost regions. The field itself just shrinks (cf. Sorensen 2004: 462-463). Frey contends that it is (metaphysically, if not nomologically) possible for the relevant structures in our brain responsible for vision to wither to such an extent that our visual experiences become spatially punctiliar. Spatially punctiliar experiences are not like isolated phosphene experiences, where the phosphene appears amongst a “sea of darkness” (Frey 2013: 75). Rather, the idea is a visual experience in which a single phenomenal point is appreciable, and nothing else whatsoever (ibid).130

What’s the upshot of smanzfeld and spatially punctiliar experiences? We know that radically different approaches to the nature of experience issue identical phenomenological predictions: one cannot tell, from introspection alone, whether representationalism, naïve

---

129 Why? Stable ganzfelds are caused by exposure to an unstructured, uniform field with degraded visual input, but the typical result in vision science is a sense of blindness, not seeing something blue, as in the case Frey talks about. For this reason I call his case a ‘smanzfeld experience’ to avoid confusion. Sorensen (2004: 462) talks about pilots experiencing a ganzfeld when flying in a homogeneously blue sky, but on my usage this is better described as a smanzfeld experience.

130 Furthermore, if spatially punctiliar experiences are possible, the dictum that visual awareness of x requires that x look extended (e.g. Pautz 2007) is false (Frey, ibid).
realism, projectivism, or sense-datum theory is true. Frey contends that the proper reaction to this is not to ditch the positive-variant of transparency altogether, but to adopt a reading of this positive variant that does not involve these unnecessary metaphysical accoutrements. And the seeds of that reading—at least to Frey—is present in our examination of things like smanzfeld and spatially punctiliar experiences: no matter how simple an experience is, an experience always involves phenomenal features being appreciated as other than oneself (ibid: 76). That is the extent to which the phenomenal character of experience is not reticent, and the kernel of truth behind the positive component of TRANSPARENCY. Harman is right to point out that we are not aware of intrinsic features of experience as intrinsic features of experience (1990: 42). But the proponent of TRANSPARENCY (not necessarily Harman) is wrong to infer the further positive claim that we experience things as having features that would ground the classification of a kind (e.g. being non-mental).131

ER, recall, claims that all objects of experience are, at a minimum, represented as being something other than oneself. This thesis was introduced as a way of delivering self-consciousness despite ELUSIVENESS. But now we have a reason to believe it independent of its

131 Frey actually gestures at ER when he says “to focus on the other is, ipso actu, to appreciate its position with respect to our phenomenal selves” (ibid: 78). Though there is certainly a convergence in our efforts, it's worth noting that Frey’s aims are different from mine; Frey is concerned with articulating an irreducibly phenomenal basis for intentionality, I am interested in developing a version of subjectivism that grants ELUSIVENESS, while using Lonergan’s insight as a springboard. In addition, ER differs crucially from his view in at least two respects. First, Frey seems to treat this sense of ‘otherness’ as a primitive, non-representational feature of experience, because it does not involve any categorical classification, and so does not determine anything like a truth-evaluable claim. On ER, this is not the case. Indeed, it shouldn’t be; to experience something as other than oneself is just as much evaluable for truth as anything else. Second, unlike Frey, the present account is not focused on intentionality as such, but on the relationship between object-consciousness and self-consciousness. Our proposal, tracking Lonergan, is in the first instance an attempt to make sense of how we can be self-conscious by being object-conscious, without the former being a species of the latter, not a theory of intentionality.

SELF-CONSCIOUSNESS DESPITE ELUSIVENESS
utility in delivering self-consciousness despite ELUSIVENESS, viz. that it seems to jibe with the phenomenology of everyday experience. And more importantly, if Frey’s assessment is right, if this sense of otherness amounts to the true positive core of the transparency phenomenon, then transparency itself confirms a central prediction of ER.\textsuperscript{132}

\textsuperscript{132} NB: This does not scuttle ER’s compatibility with HO theory. What threatens HO theory, as discussed in § 3, is a specific negative reading of transparency (‘TRANSPARENCY’), not a positive claim concerning the degree to which experience is reticent, as discussed here.
5 CONCLUSION

We have come far. Let’s recapitulate.

I have argued for three primary conclusions. The first was that the dispute over TRANSITIVITY is a verbal, substantive dispute. Upshot? To settle the dispute over TRANSITIVITY—to settle over whether a FO or HO theory accurately represents the nature of consciousness—we need in the first instance to look at the science of language. We need to look at the meaning of NAGELIAN and other ‘what it is like’-sentences.

The second was that TRANSITIVITY cannot be precisified in a way that is consistent with TRANSPARENCY. Upshot? We need to seriously rethink the role of both TRANSITIVITY and TRANSPARENCY in our theorizing about consciousness. This is important, given the pre-theoretic intuitiveness of both theses, and their purported status as truisms. For instance, if one likes TRANSPARENCY, pace Lycan (2001), there is no “simple” argument for HO theory. This is so as the supposedly “obvious” or “stipulative” premise (i.e. TRANSITIVITY) cannot be precisified in such a way that can at once support HO theory yet comport with TRANSPARENCY.

The third and final conclusion was that ELUSIVENESS puts a constraint on subjectivist theories of consciousness (it does not refute them), and that that constraint can be met without appeal to an adverbial metaphysic. The result was Egocentric Representationalism. As it stands, ER is an incomplete theory, best taken as starting advice for the subjectivist. But it is powerful. It gives the subjectivist the tools to grant arguably the most powerful objection to her view. In addition, it gives the subjectivist considerable theoretical leeway, since it can be slotted into either a FO or

133 Well, technically, my focus was W-TRANSPARENCY not TRANSPARENCY. But clearly if there is an incompatibility with the weaker former thesis, there is an incompatibility with the stronger latter.
HO framework. ER only requires a thesis about the content of conscious states, but it is silent on what makes a state conscious, and whether phenomenal character is fixed by higher-order or first-order content. All it stipulates is that such contents must involve the irreflexive relation other than, with the self as a relata. That is consistent with such contents also involving representational properties (and so higher-order content), and it is consistent with such contents excluding representational properties (and so first-order content).

We began this dissertation with four theses prevalent in the philosophy of mind: Transitivity, Nagelian, Transparency and Elusiveness. My hope is to have provided some clarity on each thesis individually, their interrelations, and their role within the broader dialectic concerning philosophical theorizing about the nature of consciousness.
CITED LITERATURE


Brogaard, Berit. (2012). What do We Say When We Say How or What We Feel? *Philosophers' Imprint* 12 (11).


Farrell, Jonathan. ‘What it is like’-Talk is not Technical Talk. Forthcoming in *Journal of Consciousness Studies*.


Hacker, P.M.S (2002). Is there anything it is like to be a bat?’ Philosophy 77 (300): 157-74.


CITED LITERATURE


York: Basic Books.


---------- The Semantics of ‘What it is like’-sentences and The Nature of Consciousness. Forthcoming in *Mind*.


Subjectivity and Selfhood: Investigating the First-Person Perspective. MIT Press.


VITA

Joseph Gottlieb  
1420 University Hall  
601 S. Morgan St.  
Chicago, IL 60607  
Email (private): joseph.gottlieb@gmail.com  
Email (institutional): jgottl2@uic.edu  
Web: joseph.gottlieb.wix.com/philosophy

EMPLOYMENT

Texas Tech University, Chicago, IL  
Department of Philosophy, Visiting Assistant Professor  
Beginning Fall 2016

University of Illinois at Chicago, Chicago, IL  
Department of Philosophy, Graduate Teaching Assistant  
August 2010 – May 2016

Fisher Investments, Woodside, CA  
Institutional Client Group, Associate  
August 2007 – July 2010

EDUCATION

University of Illinois at Chicago  
Ph.D. in Philosophy  
August 2010 – June 2016  
Dissertation: “‘Consciousness,’ Consciousness, and Self-Consciousness”  
Committee: David Hilbert (Chair), Mahrad Almotahari, David Rosenthal (CUNY), Marya Schechtman, and Daniel Stoljar (ANU)

Australian National University  
Summer 2013  
Visiting Student in Philosophy

University of Southern California  
May 2007  
B.A. in Philosophy

RESEARCH ARTICLES


TALKS (* = invited)

“Cartesian Imperativism” (with Saja Parvizian)

1. Chicago Early Modern Round Table (University of Chicago)*  
   May 2016

2. Iranian Institute for Wisdom and Philosophy (Tehran, Iran)*  
   June 2015
“Self-Consciousness Despite Elusiveness”
3. Contents of Consciousness Series (University of Manchester), unable to attend July ’15
4. UIC Department of Philosophy Work-in-Progress Series May 2015

“Presentational Character and Higher-order Thoughts”
5. Society for Philosophy and Psychology (Duke University) June 2015

“Change Blindness, Absent-Minded Perception, and Other Potentially Verbal Disputes in the Theory of Consciousness”
6. Australian National University (Canberra, Australia) August 2013

“How to Close a Gap: Explanatory Pluralism and Phenomenal Consciousness”
7. Association for the Scientific Study of Consciousness 16 (Brighton, UK) July 2012

“Good Samaritans and Parfit on Promises”
8. Illinois Philosophical Association Annual Conference (DeKalb, IL) November 2011

“Perception versus Memory: An Argument against Representationalism”
9. CUNY Graduate Center 13th Annual Graduate Conference April 2011
10. UC Boulder Rocky Mountain Graduate Philosophy Conference April 2011
11. APA Central, Great Lakes Mind and Science Consortium March 2011
12. Association for the Scientific Study of Consciousness 15 (Kyoto, Japan) June 2011

**TEACHING**

**Instructor (at UIC)**
- Introduction to Philosophy (Sp2016, Su2013, Su2012)
- Introduction to Symbolic Logic (F2015, Su2015, Su2014)
- Critical Thinking (Sp2015)
- Philosophy of Science (F2014, Sp2014, Sp2013)
- Bioethics (F2013)

**Teaching Assistant (at UIC)**
- Introduction to Symbolic Logic (F2012)
- Philosophy of Psychology (Sp2012, F2011)
- Death and Dying (Sp2011)
- Introduction to Philosophy (F2010)

**AWARDS**
- Ruth Barcan Marcus Award (Outstanding Member of the UIC Graduate Program, 2016)
- UIC Graduate College Provost's Award for Graduate Research (2012)
• Association for the Scientific Study of Consciousness Competitive Travel Award (2011)

PROFESSIONAL MEMBERSHIPS

• American Philosophical Association
• Society for Philosophy and Psychology
• Association for the Scientific Study of Consciousness