



**Consciousness Constrained:
A Commentary on *Being No One***

Josh Weisberg
CUNY Graduate Center
Philosophy and Cognitive Science
New York, New York
U.S.A.
© Josh Weisberg
jwsleep@aol.com

PSYCHE 11 (5), June 2005

KEYWORDS: Consciousness, Self-representation, Explanation, Perspectival Subjectivity, Transparency, Global availability

COMMENTARY ON: Metzinger, T. (2003) *Being No One. The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press xii + 699pp. ISBN: 0-262-13417-9.

ABSTRACT: In this commentary, I criticize Metzinger's interdisciplinary approach to fixing the explanandum of a theory of consciousness and I offer a commonsense alternative in its place. I then re-evaluate Metzinger's multi-faceted working concept of consciousness, and argue for a shift away from the notion of "global availability" and towards the notions of "perspectivalness" and "transparency." This serves to highlight the role of Metzinger's "phenomenal model of the intentionality relation" (PMIR) in explaining consciousness, and it helps to locate Metzinger's theory in relation to other naturalistic theories of consciousness. I conclude that Metzinger's theory has close affinity to "monitoring" theories of consciousness, as opposed to "first-order representational" or "global workspace" theories.

1. Introduction

Thomas Metzinger's *Being No One* is an impressive achievement. The book carefully negotiates the complex conceptual problems of consciousness, while remaining responsive to the most current empirical findings. Even those disagreeing with particular details of Metzinger's approach will have to concede that he's left very few substantial questions unaddressed. There remain a small number of opportunities, however, to draw out Metzinger on key issues, and to press him for clarification.

In my opinion, one issue that didn't get a full treatment in *Being No One* is the relationship of Metzinger's theory to rival contemporary theories of consciousness, particularly rival reductive, naturalistic theories. Metzinger contends that the various "neurological case studies" presented in the work undermine "classical theories of mind" like those of Descartes and Kant (430).¹ Deficits involving neglect or denial, for example, clearly cause problems for these venerable theories. But contemporary theories of consciousness are explicitly crafted with such cases in mind. The theories of Daniel Dennett, Bernard Baars, Fred Dretske, Michael Tye, David Rosenthal, William Lycan, and many others are arguably compatible with the presented neurological data.² Indeed, even modern-day dualists like David Chalmers take this kind of data seriously and develop their theories accordingly.³ This may seem like an amazing claim—surely *some* of the views mentioned succumb to the empirical data. Or perhaps this exposes much of consciousness studies as mere verbal disagreement?

Appealing as the suggestion may be, there is substantial disagreement between rival views. The differences turn not so much on the results of empirical studies as on the initial interpretation of the data to be explained. Each theory fixes the data according to its own proprietary set of distinctions, and this in turn delivers rival interpretations of the neurological cases. Not surprisingly, the theories in question manage to meet their own theoretical constraints, and since the empirical data is forced through the same sieve, all maintain the appearance of empirical adequacy. Therefore, if we wish to compare the success of competing views, we must find a characterization of the explanatory data that is neutral across theories. Then we will be in a position to consider the relative merits of current positions, including Metzinger's.

With that in mind, I will begin by critically evaluating Metzinger's method of fixing the explanatory data. Then I'll argue for an alternative approach—one that best isolates our commonsense conception of consciousness while avoiding the excesses of a priori reasoning rightly criticized by Metzinger. This alternative, I'll contend, offers the best route to a neutral characterization of the data. With the approach in hand, I'll reassess Metzinger's constraints and argue for a shift in focus away from "global access" and other functional- and neuronal-level concepts towards Metzinger's notions of "perspectival subjectivity" and "transparency." I'll contend that these constraints isolate a better working notion of consciousness—one both in accordance with commonsense and amenable to empirical investigation. Additionally, this shift serves to highlight the main explanatory element of Metzinger's theory: the "phenomenal model of the intentionality relation" (PMIR).

Finally, I'll (briefly) compare the recast version of Metzinger's position to rival theories. I'll argue that Metzinger's view has affinities with theories that stress the

centrality of reflective self-awareness in explaining consciousness—in particular, certain "higher-order-" and "same-order-" monitoring views.⁴ It thus arguably has distinct advantages over "first-order representation" views and "global workspace" models, as well as "bottom-up" neuroscientific approaches. In addition, by locating Metzinger's view in this region of theoretical space, it gains additional support against the conceptual worries posed by Nagel, Chalmers and others.⁵

I am in sympathy with Metzinger's overall attitude concerning the study of consciousness and I believe his theory is very much a step in the right direction. Therefore, my comments aren't intended to be destructive or undermining; instead, they offer Metzinger the opportunity to clarify and refine his position. This seems in keeping with the inclusive, open-minded outlook that pervades *Being No One*.

2. Fixing the data

In the introductory chapter of *Being No One*, Metzinger notes the importance of properly fixing the explanandum of a theory of consciousness, as well as the dangers of doing so improperly. On the one hand, a pure "bottom up" approach—one that eschews any serious analysis of what must be explained, and instead leans directly on experimental results—runs the risk of either missing phenomena of interest or of operationalizing away difficulties. On the other hand, the a priori conceptual analyses of philosophers can seem utterly disconnected from empirical science. Metzinger attempts instead to find a "middle course" that accurately characterizes the explanandum while respecting and incorporating empirical insights. We can call his approach the "method of interdisciplinary constraint satisfaction" (MICS). Metzinger considers constraints at a number of different levels of description, "constraints by which we can decide if a certain representational state is also a conscious state" (107).

The goal is to use information and techniques from a variety of sources to triangulate on a working concept of consciousness. First and foremost are the phenomenological constraints. These are based on introspective experience (110) and carry the greatest weight in fixing the data. As Metzinger notes, "maximizing phenomenological plausibility is of the highest priority for any theory of consciousness" (591). By introspectively reflecting on how things appear to us in conscious experience, we can isolate the data that must be explained. Metzinger stresses throughout that he intends to "take the phenomenon of consciousness truly seriously in all its nuances and depth" (111). He wishes to avoid charges that he has illicitly changed or avoided the subject; therefore, he gives pride of place to introspective data in his characterization of consciousness.

However, the first type of constraint must be balanced against other, more empirically-grounded constraints, to avoid "naïve-realistic assumptions and the stipulation of mysterious, nonpublic objects" (591). In addition, we need to consider the "representational," "informational-computational," functional," and "physical-neurobiological" levels of description in order to delineate our notion of consciousness. The representational level focuses on features of the intentional content of conscious experience, while the informational-computational level focuses on the computational and informational role of conscious experience. The functional level of description deals with the underlying functional profile of conscious states. The lowest-level constraints

come from neurobiology and related fields. Together, the four lower-level constraints provide a third-person check on the phenomenological constraints. Without such third-person constraints, we run the risk of characterizing consciousness in a way that rules out scientific investigation. While there may be epistemological problems surrounding the study of consciousness, it is clearly the wrong tack to define the subject as inexplicable from the outset. Moreover, Metzinger aims to make his constraints useful for the interdisciplinary study of consciousness. By explicitly incorporating concepts from a wide range of sources, Metzinger hopes to lay the groundwork for fruitful interdisciplinary research.

To illustrate MICS, I'll focus on the first defining mark of consciousness isolated by Metzinger, the "global availability" (GA) constraint. GA is largely a third-personal, functional feature of consciousness, motivated by observation of the range of behaviors open to us when our mental states are conscious. But Metzinger contends that GA also has a distinct phenomenological aspect. He claims that conscious contents are experienced as "being an integral part of a single, unified world" (120). Further, conscious experience is "characterized by flexibility, selectivity of content, and a certain degree of autonomy" (119). Being integrated into a phenomenal world provides access to a wide constellation of experienced connections; flexible, selective autonomy involves connection to a broad group of possible response states. Both these features, according to Metzinger, suggest states that are integrated with a range of processes and open to wide range of behavioral responses. This provides a link between the functionally motivated features of GA and the phenomenology of experience.

However, GA has its roots more firmly in other levels of description. Viewed in representational, informational-computational, and functional terms, GA is cashed out as information directly available to a wide array of mental systems for flexible control and quick, context-sensitive reactions. This characterization is motivated by the "global workspace theory" (GWT) championed by Bernard Baars and others, whom Metzinger cites with approval.⁶ According to Metzinger, GA in the workspace sense "is one of the very few *necessary* conditions in ascribing phenomenality to active information in a given system" (120). Finally, at the physical-neurobiological level, Metzinger discusses speculative hypotheses about the neurological realization of GA in humans. He tentatively endorses Edelman and Tononi's "dynamical core theory" as an example of the sort of approach that might be fruitful, but he concedes that little is presently known about the details of the neurological underpinnings of GA.⁷

This gives the flavor of the MICS approach. It incorporates introspection, but other sources of data are involved at the ground floor. As the GA example illustrates, much of the justification for the constraint comes from empirical work present in the literature—for example, Baars' global workspace view, which relies on a contrastive analysis between conscious and nonconscious states. This seemingly avoids the a priori biases of conceptual analysis—GA is mainly justified not by a priori intuition, but by empirical observation of the differences between conscious and nonconscious states. And, because it is grounded in empirical work, MICS avoids the "naïve" reifications of commonsense folk psychology. All told, Metzinger employs MICS to generate eleven different constraints, each analyzed at five levels of description. This yields an

enormously rich and varied characterization of the target explanandum for Metzinger's theory of consciousness.⁸

However, there are a number of problems with MICS. MICS delivers fifty-five specific elements that define the constraints. But when faced with such a wide range of characterizing features, competing theorists will likely pick and choose those elements amenable to their view, and disparage the others as extraneous or incorrect. The same unproductive cycle of debate concerning the nature of consciousness will reemerge within Metzinger's MICS structure. Metzinger may reasonably respond that this is the price of working in a complex domain, especially during the "initial preparadigmatic phase of theory formation" (213). But there is a danger from the opposite direction as well. Metzinger's MICS runs the risk of muddying the explanatory waters by including irrelevant data that doesn't belong in an initial characterization of consciousness. To the extent that we can simplify and isolate our target explanandum, we ought to do so. Letting a thousand flowers bloom may not be helpful if we are seeking to admire a specific rose.

However, consciousness surely presents an enormously complex scientific problem. Metzinger contends that "it would be methodologically naïve to assume that we can arrive at narrowly circumscribed sets of necessary and sufficient conditions for the ascription of consciousness on one level of description in the near future" (213). He worries that "consciousness" may turn out to be a cluster concept, a loose collection of related notions that fail to contain a unified structure. To best sharpen our conception, he suggests that we examine "borderline cases" culled from neuroscience. Studying these test cases "reveals implicit assumptions, helps to dissolve intuitive fallacies, and makes conceptual deficits of existing theories clearly visible" (213). But the test cases arguably are of little help. Interpreting them requires that we have a characterization of consciousness already in hand. And this isn't simply a matter of a priori modal worries. Ned Block argues on a posteriori grounds for splitting the concept of consciousness into several notions, including the distinctly phenomenological "phenomenal consciousness" and the functional "access consciousness."⁹ With this distinction in hand, he argues for interpretations of the neurological cases at odds with Metzinger's constraints. Further, there is an ongoing and widespread debate in the literature over the correctness of Block's notions. It arguably begs the question against Block to cite neurological cases as support for rival constraints when the cases themselves are at issue. Thus, we require a more neutral means of settling such disputes.

All this, I believe, points to the main shortcoming of Metzinger's MICS. MICS, especially at the lower levels of description, runs the danger of illicitly blending controversial theoretical assumptions directly into the explanandum of a theory of consciousness. For example, by employing Edelman and Tononi's "dynamical core" theory to help fix the GA constraint, MICS imports whatever assumptions those theorists use in crafting their view. But these are the very assumptions at issue. Furthermore, the dynamical core theory is a complex hypothetical model. It does not tell us per se what the data is; rather, it explains the data. To bring in the dynamic core theory at the constraint level is to build strong theoretical claims into the explanandum. This both obscures the explanatory situation and potentially begs the question against rival views that reject

Edelman and Tononi's initial assumptions. And the same goes for many of the lower-level descriptions used to pin down Metzinger's constraints.

All this serves to block useful cross-theoretic comparisons. If one theory explains consciousness characterized as GA, and another explains consciousness characterized as "what it's like for the subject," it becomes difficult to contrast the two in an illuminating way. Further, this is exactly the point where opponents of reductive, naturalistic explanations of consciousness dig in their heels and claim that their rivals employ "bait and switch" tactics. MICS fails to adequately address such worries because it imports substantial theoretical claims directly into the data without sufficient justification. Finally, MICS, by its interdisciplinary nature, runs the risk of succumbing to the very charge that Metzinger wishes to avoid: that it fails to "take the phenomenal seriously." If by "consciousness," you mean "GA," phenomenology seems to drop out of the core of the theory (I'll pursue this claim in section IV below). Further, if such a claim is supported by invoking theory-laden empirical results, it will fail to convince those who take such worries seriously. Perhaps Metzinger might argue that this is a dialectical dead end—opponents of naturalized theories will never concede that an explanation of consciousness is possible, come what may. But if we can arrive at a neutral means of fixing the data, one that doesn't beg these questions at the outset, we stand a better chance of convincing the open-minded. In the next section, I'll argue for an alternative method of fixing the data, one that better yields the "middle way" that Metzinger correctly demands.

3. Another way to fix the data

The need to properly fix the explanatory data is not unique to consciousness studies. Whenever we engage in theorizing, great care must be taken to isolate the data to be explained from the theory that does the explaining. We must especially avoid the temptation to make things easier on ourselves by defining away worrisome data. In such cases, there is little problem constructing a theory, but at the high cost of missing the issues of real explanatory interest. But the reverse worry is also present: we must not cloak our target in unnecessary mystery, even if intuition pulls in that direction. Finally, we need to avoid framing the phenomenon in terms that beg the question against rival views. Our theory will certainly seem stronger if its rivals are ruled out by definition, but this is not good scientific practice. Science moves forward when rivals compete to explain a shared set of data, not when they rule each other out by fiat. We are thus looking for a *pretheoretic* characterization of the data, one that is—as much as possible—free from the biases of particular theories.

A pretheoretic characterization is one given in everyday, commonsense terms. Such a characterization provides both a means for isolating what must be explained, as well as a condition of adequacy on theory. If a theory cannot explain why the explanandum appears as it does to commonsense—that is, if it cannot "save the commonsense appearances"—we do not count it as a successful theory. For example, we ordinarily characterize water as an odorless, colorless liquid that freezes in cold temperatures, boils when heated, etc. The claim that water is H₂O is explanatorily fruitful in large part because it allows us to understand, by connecting the everyday stuff with the concepts of chemistry, why it is that water is odorless and colorless, that it freezes and

boils at the temperatures it does, etc. The commonsense characterization thus both pins down the data to be explained, and marks out criteria for a successful explanation.¹⁰

Furthermore, if we do not employ the commonsense characterization of our explanatory target, our theoretical claims will be of little interest to those who do not agree with our alternative analysis. The claim that water is H₂O is of interest precisely because it refers to the ordinary notion of water, not a homophonic surrogate whose meaning has been stipulated by the theorizers. While it is often the case that in well-developed sciences, both the target data and the explaining theory are couched in complex technical language, this is not the case at the outset of a research program. To begin, we must have a handle on a phenomenon as we ordinarily pick it out—otherwise, we can't be sure that we've explained the features that interested us at the outset of theorizing.

This is the situation with consciousness. We need a pretheoretic characterization to get our investigation started. We are not yet at the stage where we can characterize the data in complex technical terms and still be assured that the initial elements of explanatory interest will remain intact. We therefore require a commonsense characterization of consciousness. But there are some important differences between mental phenomena like consciousness and other explanatory targets. Mental states are not observable in the way that samples of water are. We have no direct, public means of "pointing" to episodes of consciousness. Further, we have a distinct first-person way of accessing our conscious minds--our conscious mental states are introspectable. But while these facts make the situation more complex, they do not alter it fundamentally. We still want a commonsense characterization of consciousness, and that such a characterization is based in part on introspection needn't render it unfit for theorizing. Further, we can cross-check introspective folk-claims, in order to arrive at a common core. We can thus publicly triangulate on the key features of the folk conception of consciousness.

But it might be worried that this approach puts too much emphasis on a priori intuitions concerning our folk-theoretic concept of consciousness. Doesn't this method of fixing the data simply canonize naïve intuition and whatever misconceptions and confusions it contains? For example, it may seem that folk intuition holds that consciousness is nonphysical. If that's the case, what good is commonsense for pinning down the data that must be explained by a scientific theory of consciousness? But we needn't take on board everything that our commonsense conception entails. Following the lead of Daniel Dennett, we can allow that the folk view is authoritative on how things *seem* to us in introspection, but not about our mental states' underlying nature.¹¹ Thus, our commonsense view pins down the appearances that a theory of consciousness must explain, but it remains neutral as to how those appearances are realized. And this is arguably in keeping with the folk view in any event. While it may be the case that consciousness doesn't introspectively appear to be a physically realized phenomenon, it isn't therefore the case that it appears nonphysical.¹²

Furthermore, while armchair thought experiments may at times be useful in delineating the folk characterization of consciousness, they are not definitive of the phenomenon and must be used with care. Our ordinary folk conception is at home in normal, everyday situations. If we take that conception out of the realm of the ordinary and off to distant possible worlds, it will likely be stretched beyond the breaking point.

Therefore, our modal intuitions about possible cases are of limited value in fixing the data. The folk conception is just not refined enough to provide clear results concerning the application of our concepts in widely-divergent counterfactual situations. Can we really be sure how we'd apply our concept of consciousness if the laws of nature were different? This is not to say that hypothetical cases have no role to play, but it is to insist that such cases be tied relatively closely to reality, and to caution against relying too heavily on modal intuitions.

What's more, the process of laying out our commonsense characterization of consciousness is best seen as a posteriori. The need to cross-check and triangulate on the core of the folk conception insures that no one person's armchair reflection is enough to pin down just what we ordinarily mean by "consciousness." And even if this process can be usefully informed by self-reflection, that doesn't entail anything about the underlying nature of consciousness. Since we are all competent users of folk psychology, we have a degree of access to the meaning of our mental state terms, including the ones picking out conscious experience. But that doesn't mean that we thereby know a priori that the mind is nonphysical—or physical, for that matter. But what we want explained, the explanandum that our theory must answer to, is provided by the folk characterization of consciousness, consciousness as it is ordinarily picked out by the folk, in everyday situations, including those involving introspective reflection.

It might be worried that this method bears too close a resemblance to Daniel Dennett's "heterophenomenology," and thereby fails to "take the phenomenal seriously."¹³ Dennett proposes taking folk psychology as providing a "theorist's fiction," one that lays out a phenomenal world but makes not commitment to its reality. But this charge is misplaced. Just because Dennett's positive theory has seemed to some critics to sell the phenomenal side of the mind short does not entail that his initial method of fixing the data is inadequate. That Dennett often opts to explain away—rather than to explain—the folk appearances does not mean he hasn't accurately pinned down how things seem from the normal first-person point of view. In fact, when looked at in isolation, and ignoring issues of a priority, Dennett's approach to data fixing is very similar to that employed by Frank Jackson and David Chalmers.¹⁴ Jackson and Chalmers also turn to the folk conception to fix the data, and both contend that a pretheoretic commonsense characterization of the consciousness is a crucial first step to theorizing. So long as we are careful not to undersell the phenomenal elements of the folk characterization of consciousness, we'll avoid the danger of failing to take the phenomenal seriously. If Jackson and Chalmers believe the method is adequate to fix the data, there is little worry that it leaves out the phenomenal side of mind.

Therefore, we should attempt to isolate our commonsense characterization of consciousness. This provides us with the best chance to arrive at a neutral characterization of the data, one that doesn't beg important theoretical questions at the outset. We should stick as best we can to everyday cases in order to triangulate on the core of the folk conception. And we should remain appropriately noncommittal concerning the underlying nature of the phenomenon. By sticking to these guidelines, we can arrive at a characterization that both takes the phenomenal seriously, but does not a priori rule out the possibility of meaningful inquiry. However, it follows that if a purported feature of consciousness can't be captured in commonsense terms, it is suspect

as an initial characterization of the data that a theory must explain. With this in mind, I'll turn back to Metzinger's constraints and isolate the subset that best conforms to our commonsense characterization of consciousness.

4. Global Availability Rejected

With our alternative method of fixing the data in hand, we can re-evaluate Metzinger's constraints. In this section, I'll argue that the GA constraint should be jettisoned altogether and instead that a better working concept of consciousness can be drawn from Metzinger's "perspectivalness" and "transparency" constraints. To begin, I'll contend that GA does not pick out consciousness as it is characterized by commonsense.

To add some additional detail concerning the nature of the GA constraint, Metzinger writes that:

Phenomenally represented information is exactly that subset of currently active information in the system which possesses one or more of the following three dispositional properties:

- availability for *guided attention* (i.e., availability for introspection; for nonconceptual mental metarepresentation);
- availability for *cognitive processing* (i.e., availability for thought; i.e., for mental concept formation);
- availability for *behavioral control* (i.e., availability for motor selection; volitional availability) (31).

Support for these claims comes from the phenomenological and lower-level descriptions cataloged in section II above. Phenomenologically, according to Metzinger, GA is experienced as flexible, selective, autonomous processing and the presence of a unified phenomenal world. At the lower levels, GA gets its support from the theoretical work of Baars, Edelman and Tononi, and others—GA is posited in those theories as the informational/computational and functional difference between conscious and nonconscious states. But to what extent does GA gain support from our commonsense conception of consciousness?

First of all, in and of itself, theoretical support from ongoing research is arguably irrelevant to a commonsense justification. Such research is not cast in folk terms, and it must presuppose some version of the distinctions at issue. But we can consider if the aspects of GA described in the above quote are sufficient for a mental state to be conscious, from the commonsense point of view. Working in reverse order, from availability for behavioral control, there are many ordinary cases where beliefs, desires, emotions, and even perceptual states actively influence our behavior—even behavior intuitively labelled "volitional"—despite the fact that those states aren't conscious in any intuitive sense. An unconscious belief that there is beer in the fridge may move me into the kitchen, or an unconscious desire to watch the game may place me on the couch with the remote in my hand. While these sorts of behaviors often occur consciously, they also occur without subjects being in any way conscious of the beliefs and desires that drove

their actions. There is nothing contradictory, incoherent, or even particularly odd about describing things in such terms.

Likewise, my jealousy may not be conscious, but it may motivate rude behavior towards a friend of my wife. I may vigorously deny that I'm jealous, though later events may conclusively prove me wrong. Whole ranges of behavior can be instigated in this manner, requiring that the unconscious state be accessible to behavior-causing systems. Metzinger might argue that such behavior lacks the requisite flexibility to be labeled globally available. But we engage in a wide range of subtle and sophisticated behavior when we're involved in intense personal relationships, and not all of this behavior is engaged in consciously. The challenge for Metzinger is to effectively delineate those behaviors that exhibit the requisite flexibility from those that do not.¹⁵ Finally, even perceptual contents may show the profile of GA despite occurring nonconsciously, from the commonsense point of view. For example, I may walk down the street deep in thought and still arrive at my destination safely. Intuitively, not everything registered by my senses is conscious. I don't have any experience of many of the things I pass; there is no phenomenology vis-à-vis those features that I'm aware of. But such percepts must be available to control my action if I am able to arrive without injury. Walking through the streets of Brooklyn while contemplating philosophy requires complex, spontaneous reactions, and novel patterns of motor action. But it does not intuitively entail that all the guiding perceptions must be conscious. All these examples indicate that we can possess active representational contents available for behavioral control even when such contents fail to be conscious in any manner recognized by commonsense. This might seem to beg the question—this interpretation isn't forced on us, surely. But the presented cases have clear, coherent folk-psychological readings in terms of nonconscious states. What reason do we have to deny these interpretations? That they fail to fit with our favored theory of consciousness, of course, begs the question in the opposite direction. Given that commonsense is a neutral starting point, rival interpretations arguably have the burden of proof here.

A similar case can be made concerning availability for cognitive processing. As a wide range of examples in cognitive science makes apparent, subliminal or masked primed stimulus can influence our cognitive processes. And the easy folk-comprehension of such cases points to a coherent folk-psychological explanation in terms of nonconscious states. Furthermore, most of us have had the experience of struggling with a complex problem only to arrive at the solution when we stop consciously attending to it. In such "eureka" cases, a coherent commonsense explanation holds that we continue to process the problem nonconsciously. But in order for the problem to be solved, the states involved in the puzzle have to be available to cognitive processes. This description fits well with our ordinary categorization of such cases. Relevantly, people do not feel compelled to say that our mental states must have been conscious because we accessed them in problem solving.

Finally, availability to introspection is plausibly related to our commonsense characterization of consciousness, but not in the constitutive manner implied by Metzinger's constraint. States are not conscious because they are introspectable; they are introspectable because they are conscious. And this doesn't provide us with an illuminating link between consciousness and introspection. Intuitively, we could have

metacognitively-inaccessible conscious states. Further, it's often argued on the basis of experimental results that small children are unable to introspect, or are limited in their introspective abilities. Does it follow that these children are thereby unconscious? Arguably, there is nothing in our folk conception that compels such a reading, and it does feel counterintuitive. This suggests that access to introspection isn't constitutive of consciousness, but rather that there is something about conscious states that renders them accessible to introspection.

Taken together, we can conclude that GA is not sufficient for consciousness, from the commonsense point of view. Metzinger himself notes that without the addition of "perspectivalness" and "transparency" (see below) GA states are only "minimally conscious" at best (204-5). "Minimally conscious" states are not subjective; they are not *for* any subject at all. Arguably, this does not answer to any commonsense notion of consciousness. If mental states aren't relevantly *for* a subject, they will not intuitively be conscious states. States that don't appear to anyone are not part of our everyday notion of consciousness. Thus, GA isn't sufficient for consciousness.

But Metzinger is only arguing for GA as a *necessary* condition for consciousness. Why should it worry him that it isn't sufficient? First off, if it isn't sufficient, we must enquire what needs to be added to make it sufficient. This investigation, I'll argue below, suggests that GA isn't necessary for consciousness, once we fill in the other conditions. In fact, GA is better construed as a functional property of representational states, independent of any useful notion of consciousness. Further, we can plausibly explain GA in consciousness as a derivative feature of other processes, rendering it irrelevant to fixing the data. But let us briefly consider if all conscious states really have the *phenomenological* features of GA identified by Metzinger. Metzinger writes, "[GA] is an all-pervasive functional feature of my conscious contents, which *itself* I... subjectively experience... as my own flexibility and autonomy in dealing with these contents" (118). But surely there are a wide range of conscious contents for which I do not experience this feeling of flexibility and autonomy. If an anvil drops on my foot, I will consciously feel it, in all likelihood. But I will have no control over how I react to the experience, and no flexibility in what I do next, given my conscious state. I will scream in pain, and I will not be able to continue thinking about philosophy, football, or whatever. But the state will most definitely be conscious. It seems to me that such examples are numerous, indicating that at least from the phenomenological point of view, GA as glossed by Metzinger is not necessary for consciousness.

There is an additional phenomenological mark of GA described by Metzinger, the sense of "being in the world" present in consciousness. This is further unpacked as an independent constraint, which Metzinger calls "globality" (131-143). However, the experience we have as conscious creatures of being in the world is arguably unrelated to any sort of functional access relation. Our conscious experience presents us with a rich, ever-shifting manifold of stimuli, and we as subjects are situated in relation to these experienced objects. But why should presence in this manifold, by itself, offer any sort of functional advantage, particularly in terms of increased availability of contents? Why wouldn't integration instead reduce the salience of stimuli by blending them in with the "buzzing, blooming confusion" of an experienced world? While it may follow as a result of a theory that conscious states gain such a functional property, it is not part of an initial

pretheoretic characterization of the phenomenon. Further, I will argue below that the phenomenal aspects of globality follow from a theoretical explanation of more basic characterizing constraints of consciousness. To be more specific, when perspectivalness and transparency are explained in terms of a PMIR, globality falls out as a byproduct. Thus, globality, as a phenomenological aspect of GA, is not definitive of consciousness, but something that results from the presence of the mechanisms that explain consciousness, defined in independent terms.

I conclude that GA is not a part of a commonsense characterization of consciousness. It is not sufficient in any intuitive sense, and it likely isn't necessary, when viewed pretheoretically. While GA may accompany conscious states, and may result from the processes that in fact make us conscious, it is not part of the data that a theory of consciousness must explain. It seems better situated in the computational-informational and functional camps. It is most likely a great benefit to liberate the study of GA from the shrill debates of consciousness studies. To the extent that a phenomenon can be explained in terms of nonconscious processes, it can be explained free from worries of "hard problems" and "explanatory gaps."

5. Consciousness as Transparent Perspectivalness

Instead of relying on GA to fix the data, I propose that we focus on Metzinger's "perspectivalness" and "transparency" constraints. Perspectivalness, according to Metzinger, is rooted in the fact that conscious experience is tied to an individual perspective (156). Perspectivalness means that in conscious experience we have a point of view, that we are a focus of experience (157). Without perspectivalness, consciousness is not *subjective*. Transparency, in Metzinger's sense, occurs when we cannot introspectively access the processing stages that lead to the formation of our conscious states. Because we aren't aware that we are the source of our conscious states, they seem to occur spontaneously. This results in the apparent *immediacy* of conscious experience. Metzinger writes,

What is inaccessible to conscious experience is the simple fact of this experience taking place in a *medium*. Therefore, transparency of phenomenal content leads to a further characteristic of conscious experience, namely, the subjective impression of immediacy (169-170).

When an experience is transparent in this sense, we are unaware of any mediation between the experience and our access to it. It is directly present to us.

I contend that taken together, transparency and perspectivalness form a well-justified working concept of consciousness. States that possess the property of perspectivalness are properly *for* a subject, they are states that the subject is aware of in a special way. And transparent states are accessed with subjective immediacy, with the characteristic directness of conscious experience. We don't consciously infer or reason to our conscious states; they just happen, and we are seemingly in direct contact with them. As I argued above, states that are in no way *for* a subject--states that float free of any connection to an experiencing agent—are not conscious states in any intuitive, folk-

psychological sense. "Conscious experience" and "subjective experience" are synonymous from a commonsense perspective. Further, conscious states are *for* us in an apparently direct way, without the appearance of any mediation. I conclude that if states are *transparently perspectival*, they are conscious, from a commonsense perspective. This provides us with a neutral, pretheoretic characterization of the data.

This characterization maps well onto what David Rosenthal has termed the "transitivity principle" (TP), the idea that *conscious states are states that we are conscious of ourselves as being in*.¹⁶ If we are in no way conscious of ourselves as being in a state, that state is not intuitively conscious. Rosenthal also argues that the awareness we have of our conscious states must seem, from the subject's point of view, to be immediate. Thus the TP captures the same features as those picked out by transparent perspectivalness. This provides an independent line of support for the characterization, and it also helps to fend off charges that the data is being misdescribed or altered. Additionally, Rosenthal (2004) and Eric Lormand (forthcoming) independently argue that the TP is the best reading of Nagel's claim that an organism is in conscious states when "there is something it is like to *be* that organism--something it is like *for* the organism" (1974: 519, emphasis in original). Rosenthal and Lormand both argue that if we are not conscious of our states, they are not intuitively for us. Nagel's formulation is often cited in the literature as the most effective characterization of the phenomenal nature of conscious states. Further, it's argued that such states defy functional or even physical explanation. But if Nagel's formulation is best paraphrased in terms of the TP, there is no such worry. States that we are aware of ourselves as being in are plausibly explainable in representational or functional terms.

Metzinger argues, however, that neither transparency nor perspectivalness is necessary for consciousness. He contends that some conscious representations are "opaque" because we can introspectively access that such states are internal representations. He cites propositional attitude states as a prime example. We can be aware, he contends, that such states are representations of our own creation. But this seems to miss the phenomenon at issue. We may be aware that our belief that snow is white is a representation, but that does not entail that we can introspectively access the *processes* that led to its presence in our conscious experience. We may even come someday to know exactly how such processes work, and thus be aware, in a sense, of the processes that must have occurred when a belief is conscious. But this still falls short of seemingly direct introspective access to such processes. And this is arguably what matters for transparency. Our states seem immediately accessed because we aren't *introspectively* aware of the processes that led to their presence. Even if we know in some other way how they are formed or that they are representations, that won't undermine transparency. Thus, I believe that transparency is a necessary feature of consciousness.¹⁷

This suggests an ambiguity in Metzinger's use of "transparency." In places he uses it to mean that we can't introspectively access the processes that account for our conscious representations. This is the sense that I have been stressing and that I believe is partially constitutive of consciousness. However, sometimes he uses it to mean that we can't cognitively recognize that our representations are representations rather than nonmental objects. This clearly sometimes fails for conscious representations without

altering their conscious status. However, this doesn't seem to be the sort of awareness relevant to the immediacy of consciousness.

Metzinger also argues that there can be states of *nonsubjective* consciousness (157). He cites examples of religious experiences or states present in certain neurological disorders, like "akinetetic mutism." However, I do not believe he properly describes such cases. The religious experience of selflessness is better described as a complete lack of awareness of the self as such in consciousness, not as an example of consciousness that is in no way for a subject. We may represent ourselves as indistinct from the rest of the world or we may represent ourselves under an extremely unadorned self-concept. But we are still arguably *aware of ourselves* as a being indistinct from the world or being "empty" or "formless." These states are still for us, though we are no longer conscious of ourselves in the ordinary way. This response may seem unsatisfying because it simply restates my thesis. However, I question whether Metzinger has clearly established his reading of these cases. I also think these cases are particularly unclear, suggesting the vague questions "what is it like to be a yogi" or "what is it like to be an akinetic mute." Such borderline cases will probably remain unsettled until we have a well-entrenched theory of consciousness. For the time being, I will press on with my characterization of consciousness and explore how Metzinger's PMIR provides a good explanation of transparent perspectivalness.

Conscious states are states that are transparently perspectival. This means that they are states that we are immediately conscious of ourselves as being in. This characterization highlights what I take to be the most plausible of Metzinger's constraints, and it captures our folk sense of consciousness. It thus provides a neutral characterization of the data, one that is phenomenologically accurate and as free as possible of question-begging assumptions. It also arguably offers a satisfying explication of Nagel's "states there is something it is like for the subject" locution. If a theory successfully explains this explanandum, it stands as a successful theory of consciousness. The main tool used by Metzinger to explain consciousness in this sense is his "phenomenal model of the intentionality relation" or PMIR. The PMIR "is a conscious mental model and its content is an ongoing, episodic *subject-object relation*." It depicts "a certain *relationship* as currently holding between the system, as transparently represented to itself, and an object component" (411). For example, a PMIR represents to the effect that "I am someone who is currently visually attending to the color of the book in my hands" or "I am someone now deciding to get up and get some more juice" (Ibid.). There is a subject component, made up of a representation of the self,¹⁸ an object component that can take any number of representational contents, including sensory contents, intentional contents, or even models of the self, and a representation of the relation between the two, cast in broadly intentional terms.

The PMIR represents the self as being in certain states. It thereby makes us aware of those states—awareness is a relationship plausibly cashed out in representational terms. Further, the PMIR explains why our conscious states appear unmediated: the PMIR does not represent the processes that underlie or lead to our conscious states. It only delivers what is represented in the subject and object components and their relation. What is represented thereby appears to us directly. Thus, the PMIR accounts for

transparent perspectivalness--in other words, it accounts for consciousness in the sense defended here.¹⁹

Metzinger presents a number of other constraints, but it is beyond the scope of this commentary to address them directly. It is enough to point out that they are not constitutive of consciousness, but instead constrain the workings of the PMIR. They characterize important features that PMIR representations must convey: the qualitative character of consciousness ("convolved holism," "dynamicity," "sensory intensities" and "homogeneity"), its temporal nature ("activation in a window of presence"), and its functional profile ("offline simulation" and "adaptivity"²⁰). This helps isolate those constraints that pick out consciousness itself from those that determine how particular conscious contents are experienced. Looking at these constraints in light of the PMIR also points to the direction of further research. We should theoretically investigate representations having the capacity—when slotted into a PMIR—to make us conscious of the rich, variable features of conscious experience.

A last word about the GA constraint. I mentioned that considering the workings of the PMIR gives us reason to doubt that GA is a necessary condition for consciousness. We can see this by noting how the PMIR accounts for the phenomenal aspect of GA, "globality." Metzinger writes that the PMIR "consists of a transparent subject component, and varying object components... transiently being integrated into an overarching, comprehensive representation of the system as standing in a specific relation to a certain part of the world" (413). It delivers the "being in a world" phenomenology of globality by representing the subject as being thus integrated. While globality may always accompany consciousness, this does not indicate that the functional GA does as well.

We can also consider how being integrated into a PMIR might add to the informational-computational or functional profile of a representational state, suggesting that GA might be a byproduct of the way we instantiate transparent perspectivalness, rather than a necessary condition for consciousness in its own right. One possibility is that the PMIR reinforces or strengthens the contents that it incorporates, so that they achieve "cerebral celebrity" and become available to a wide array of systems. Alternatively, the PMIR might actively re-represent lower-level contents instead of simply incorporating them into its structure. This would "double-up" a representation's content, plausibly upgrading its functional profile. Finally, the PMIR might represent in a format that makes its contents globally available. It may be that the PMIR represents in a more efficient or broadly accepted format, thus aiding connectivity. Or perhaps it interacts more effectively with higher cognitive functions or language, bringing powerful recursive structures to bear on the contents of the PMIR. It's clearly an empirical question, but I hope I have indicated how the PMIR may achieve GA. Thus GA may be a by-product—rather than a definitive mark—of consciousness.

To conclude, my revised reading of Metzinger's theory strengthens it in several ways. First, it isolates a more focused characterization of consciousness, allowing many of the processes and structures presented by Metzinger to be dealt with on the nonconscious level. Second, a characterization of consciousness in terms of transparent perspectivalness fits with our commonsense conception of consciousness, thus delivering a neutral fixing of the data, and one relatively immune from a priori, modal attack. Third, it further highlights the crucial role of the PMIR in Metzinger's theory--not that

Metzinger in any way fails to recognize the centrality of this structure, but on this reading, it is freed from various forms of unnecessary theoretical baggage. Finally, it allows for a re-orienting of the additional constraints in terms of how they fit with the PMIR. To reiterate my introductory remark: I am not offering any knock-down objections to Metzinger. I feel his view is very much on the right track, especially once it's tweaked in the ways I've suggested. Rather, I am curious to hear why Metzinger would reject my suggestions, and eager to hear more about how he justifies MICS and the particular constraints—especially GA—it delivers.²¹

6. Conclusion

At the beginning of this commentary, I promised to consider how Metzinger's view stacks up against rival contemporary theories. Instead, I've spent my time arguing about how to pin down the preliminary data a theory of consciousness must explain. This, I think, isn't unreasonable, in part because we are at an early stage in formulating empirical theories of consciousness, and in part because consciousness by its nature invites a wealth of methodological and conceptual confusions. In any event, given the characterization of the data that I've argued for, we can finally offer some rough and ready comparisons between Metzinger's view and its rivals.

As noted above, when the data is framed in terms of transparent perspectivalness, Metzinger's view shows a strong affinity with the class of theories that take a form of self-consciousness to be at the heart of an explanation of conscious experience. Those views embrace—to a greater or lesser degree—the idea that conscious states are states we are conscious of ourselves as being in. This includes the higher-order-perception views of Armstrong and Lycan, Rosenthal's higher-order-thought hypothesis, Rocco Gennaro's wide-intrinsicality view, Robert Van Gulick's higher-order-global-states theory, and Uriah Kriegel's same-order-monitoring view. (HOP, HOT, WIV, HOGS, and SOMT, respectively, for those who like alphabet soup.) Future work in this area should attempt to refine and differentiate these models, with the goal of generating divergent, empirically testable predictions.²²

Looking at competing theories, if the data is fixed in the manner I've argued for, arguably Metzinger's view has theoretical advantages over both GWTs, and "first-order representational" (FOR) views like those of Tye and Dretske. GWTs generally fail to explain how we are conscious of our conscious states, or, to put it another way, how mere access turns into full-fledged perspectivalness. They are also saddled with the problem of motivating GA as a pretheoretic mark of consciousness and of warding-off legitimate charges of subject-changing. Thus Metzinger's view is a substantial improvement over simpler GWT models, both because of its richer theoretical structures and its potential to address the proper explanandum.

FOR views likewise fail to explain perspectivalness, given that all they have to work with are intentional representations of the world. While both Tye and Dretske appeal to the additional idea that conscious representations are poised to influence our conceptual systems, they fail to explain how this accounts for transparent perspectival access to our conscious states. Metzinger's theory again has a much richer set of representational tools at its disposal, particularly in terms of the PMIR, and it is oriented towards the right explanandum. In a sense, Metzinger's view properly explains the nature

of GA, and the availability of conscious representations to the conceptual system. Thus it satisfies the desiderata of these rival views, while offering a more satisfying account of the central datum any theory must explain. It also should be noted that Metzinger's theory is an improvement on pure bottom-up approaches, because it properly attends to fixing the data a theory must explain—modulo my "friendly amendments," of course!—and because it seriously addresses the phenomenology of conscious experience. Further, the impressive detail of the view allows for the easy incorporation of details culled from lower-level investigations. All told, the well-developed theory presented in *Being No One* compares favorably to its contemporary rivals. It represents an important step on the road to a scientific explanation of consciousness.²³

Notes

1. All references to Metzinger's work are to *Being No One: The Self-Model Theory of Subjectivity*, Cambridge, MA: MIT/Bradford Books, 2003. All emphases are in the original.
2. Dennett, 1991; Baars, 1988, 1997. Dennett and Baars defend versions of "global access" or "workspace" hypotheses. Dretske, 1995; Tye, 1995, 2000. Dretske and Tye defend "first-order representational" theories. Rosenthal, 1997, 2002; Lycan, 1987, 1996. Rosenthal and Lycan defend "higher-order representational" theories. See also Armstrong, 1968, 1980; Carruthers, 2000; Papineau, 2002; etc.
3. Chalmers, 1996.
4. See especially the "same order" views of Gennaro and Van Gulick. Both stress the idea that consciousness involves self-awareness. Rosenthal's higher-order theory also contends that consciousness states are ones we are *conscious of ourselves* as being in. Armstrong, 1968a, 1980; Rosenthal, 1986, 1997, 2002; and Lycan, 1987, 1996 defend "higher-order" views. Gennaro, 1996, 2004; Van Gulick, 2001, 2004; and Kriegel, 2003, forthcoming defend "same-order" views.
5. Nagel, 1974; Chalmers, 1996; McGinn, 1989; Levine, 2001.
6. Baars, 1988, 1997; Dehaene and Naccache, 2001; Dennett, 1991, etc. See Dennett, 2001 for a philosophical overview.
7. Edelman and Tononi, 2000.
8. The constraints are "global availability," "activation in a window of presence," "Integration into a coherent global state," "convolved holism," "dynamicity," "perspectivalness," "transparency," "offline activation," "representation of intensities," "homogeneity," and "adaptivity." See chapter 4 of *Being No One* for a detailed exposition of these constraints. Due to limitations of space, in this commentary I focus mainly on global availability, perspectivalness, and transparency.
9. Block, 1995, 2001, 2005.
10. It is important to stress that this is not to reject the idea that our commonsense view of the mind is itself a sort of theory, nor is it to claim that folk psychology presents no biases into our characterization of consciousness. Instead, it is to stress that between the sorts of

theories considered here—Metzinger's in particular, and other reductive rivals—commonsense provides a neutral ground. All such theories claim to explain the appearances, or at least explain them away. But this requires pinning down the appearances that need saving in the first place; commonsense, I am arguing, is the place to do so. Thanks to an anonymous reviewer for pressing this issue.

11. See Dennett, 1991, chapter 4.

12. See Armstrong, 1968b.

13. Dennett, 1991, chapter 4. For criticisms see, for example, Block, 1993.

14. Jackson, 1998; Chalmers, 1996, chapter 1; Chalmers and Jackson, 2001.

15. Thanks to Tim Bayne for pressing this point.

16. Rosenthal, 2000. See also Rosenthal, 1986, 1997, 2002; Armstrong, 1968, 1980; and Lycan, 1996, 2001.

17. Metzinger seemingly endorses this claim. In describing a system that satisfies the globality, presence, and perspectival constraints but fails to satisfy the transparency constraint, he concludes, "The present theory would have to describe it as unconscious because the transparency constraint was interpreted as a necessary condition in the conceptual ascription of phenomenality..." (614). Since he is here discussing various conceivable creatures, he may not be endorsing transparency as a necessary condition tout court. But the quote is suggestive.

18. Metzinger of course develops a detailed theory of the self (or its "successor concept"), especially in chapter 6 of *Being No One*. He explains the self in terms of a "phenomenal self model" (PSM). I will not address the specifics of that theory here.

19. Metzinger appears to endorse the spirit of this claim, writing

The existence of the PMIR is what generates full-blown consciousness, and it is precisely this feature of the deep representational structure of our conscious model of reality which appears as of highest relevance to most of us. Full-blown conscious experience is more than the existence of a conscious self, and it is more than the mere presence of a world. It results from the dynamic interplay between this self and the world, in a lived, embodied present (417).

Metzinger is willing to countenance other, more "minimal" forms of consciousness, but I believe that the "full-blown" form is the proper target of a theory of consciousness, and that the other levels are better classified as nonconscious.

20. I'll leave aside discussion of the evolutionary function of consciousness here. I think this is a wide open question, and it may be that many of the benefits of consciousness are useful "spandrels" rather than full-blown adaptations. Still, to the extent that the PMIR is useful, it accounts for the evolution of consciousness.

21. I have one final phenomenological concern with Metzinger's view. He makes too much of the system phenomenal. The PMIR, for example, has a rich and complex representational profile. But it is the results of the PMIR that are conscious, not the structure itself. We are generally conscious of the represented world in consciousness, not a range of complex representational structures. I believe, therefore, that the model is

better construed as a NMIR, a *nonconscious* model of the intentionality relation. This also helps explain the transparency of consciousness: we aren't consciously aware of this complex representational structure, only its results. And the same goes for the PSM. A phenomenal self model only becomes conscious when actively integrated into a NMIR, and then only in the transparent way that the NMIR represents it. Furthermore, if GA is no longer a mark of consciousness, it's not clear why such structures would be conscious. It is only if they themselves are picked up as either the object or subject component of an NMIR that they'd be conscious. In fact, it presents a potential regress if the MIR is a PMIR and GA isn't available. The highest-level MIR must be nonconscious to block the regress. Not only, then, does an NMIR give a better account of ordinary phenomenology, but it is forced on us if we jettison GA as definitive of consciousness. Things become less cluttered phenomenologically and more consistent theoretically.

22. Metzinger argues that the awareness involved in consciousness is nonconceptual in nature. In support of this claim he cites work by Diana Raffman concerning our ability to detect many more color differences than we can explicitly recognize and name (1995). This puts his view more in line with the HOP theories of Lycan and Armstrong and the HOGs view of Van Gulick. However, I argue in Weisberg, 2003 that Raffman's results are inconclusive on this point (2003: 92). Thus there is still arguably important work to do in differentiating the models, even those involving conceptual awareness. Thanks to an anonymous reviewer for pressing this point.

23. Thanks to Dorothée Legrand, Tim Bayne, and an anonymous reviewer for PSYCHE for helpful comments.

References

Armstrong, D. M. (1968a). *A Materialist Theory of the Mind*. New York: Humanities Press.

Armstrong, D. M. (1968b). "The headless woman and the defense of materialism." *Analysis* 22, 48-49.

Armstrong, D. M. (1980). *The Nature of Mind and Other Essays*. Ithaca: Cornell University Press.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Baars, B. J. (1997). *In the Theater of Consciousness: The Workspace of the Mind*. Oxford: Oxford University Press.

Block, N. J. (1993). Review of D.C. Dennett, *Consciousness Explained*. *Journal of Philosophy* 90: 181-93.

Block, N. J. (1995). "On a Confusion about a Function of Consciousness." *Behavioral and Brain Sciences* 18, 2. In N. Block, O. Flanagan, and G. Güzelde, eds., (1997). *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: The MIT Press, 375-416.

- Block, N. J. (2001). "Paradox and Cross Purposes in Recent Work on Consciousness." *Cognition*, 79, 197-219.
- Block, N. J. (2005). "Two Neural Correlates of Consciousness." *Trends in Cognitive Science*, Vol. (9), 2.
- Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.
- Chalmers, D. J. (1996), *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.
- Chalmers D. J. and Jackson, F. (2001) "Conceptual Analysis and Reductive Explanation," *Philosophical Review* 110: 315-361.
- Dehaene, S. and Naccache, L. (2001), "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework" *Cognition* 79, 1-37.
- Dennett, D. C. (1991), *Consciousness Explained*, Boston: Little Brown.
- Dennett, D. C. (2001), "Are We Explaining Consciousness Yet?" *Cognition* 79, 221-237.
- Dretske, F. (1995), *Naturalizing the Mind*, Cambridge: The MIT Press.
- Edelman, G. and Tononi, G. (2000), *A Universe of Consciousness: How Matter Becomes Imagination*, New York: Basic Books.
- Gennaro, R. J. (1996), *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*, Amsterdam and Philadelphia: John Benjamins Publishers.
- Gennaro, R. J. (2004), "Higher-Order Thoughts, Animal Consciousness, and Misrepresentation: A Reply to Carruthers and Levine," in R. J. Gennaro, ed., *Higher-Order Theories of Consciousness*, Amsterdam and Philadelphia: John Benjamins Publishers, 45-66.
- Jackson, F. (1998), *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Clarendon.
- Kriegel, U. (2003), "Consciousness as Intransitive Self-Consciousness: Two Views and an Argument," *Canadian Journal of Philosophy* 33: 103-132.
- Kriegel, U. (forthcoming), "The Same-Order Monitoring Theory of Consciousness," in *Consciousness and Self-Reference*, U. Kriegel and K. Williford, eds., Cambridge MA: MIT Press/Bradford Books.
- Levine, J. (2001), *Purple Haze: The Puzzle of Consciousness*, New York: Oxford University Press.
- Lewis, D. K. (1972), "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy*, L, 3 (December), 249-258, reprinted in Lewis, D. *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 248-261.
- Lewis, D. K. (1994), "Reduction of Mind," reprinted in Lewis, D. *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 291-324.

Lormand, E. (forthcoming), "The Explanatory Stop-Gap," *Philosophical Review*, <http://www-personal.umich.edu/~lormand/phil/cons/stopgap.htm>.

Lycan, W. G. (1987), *Consciousness*, Cambridge, MA: The MIT Press.

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

Lycan, W. G. (2001), "A simple argument for the higher-order representation theory of consciousness," *Analysis* 61.1: 3-4.

McGinn, C. (1989), "Can We Solve the Mind-Body Problem," *Mind*, 98:891, reprinted in Block, N., Flanagan, O., and Güzeldere, G., (eds.), (1997), *The Nature of Consciousness: Philosophical Debates*, The MIT Press, Cambridge, MA, 529-542.

Metzinger, T. (2003), *Being No One: The Self-Model Theory of Subjectivity*, Cambridge, MA: MIT/Bradford Books.

Nagel, T. (1974), "What is it like to be a Bat," *Philosophical Review*, 83, reprinted in N. Block, O. Flanagan, and G. Güzeldere, eds., (1997), *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: The MIT Press, 519-527.

Papineau, D. (2002), *Thinking About Consciousness*, Oxford: Clarendon.

Quine, W. V. O. (1951), "Two Dogmas of Empiricism" *Philosophical Review*, 60:20-43.

Quine, W. V. O. (1960), *Word and Object*, Cambridge, MA: MIT Press.

Raffman, D. (1995), "On the persistence of Phenomenology" in T. Metzinger ed. (1995), *Conscious Experience*, Thorverton, UK: Imprint Academic.

Rosenthal, D. M. (1986), "Two Concepts of Consciousness," *Philosophical Studies* 49, 3 (May): 329-359.

Rosenthal, D. M. (1997). "A Theory of Consciousness," in N. Block, O. Flanagan, & G. Güzeldere, eds., *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: The MIT Press, 729-753.

Rosenthal, D. M. (2000), "Consciousness and Metacognition," in D. Sperber, ed., *Metarepresentation: Proceedings of the Tenth Vancouver Cognitive Science Conference*, New York: Oxford University Press.

Rosenthal, D. M. (2002), "Explaining Consciousness," in D. Chalmers, ed., *Philosophy of Mind: Contemporary Readings*, Oxford: Oxford University Press.

Rosenthal, D. M. (2004), "Varieties of Higher-Order Theory," in R. J. Gennaro, ed., *Higher-Order Theories of Consciousness*, Amsterdam and Philadelphia: John Benjamins Publishers, 19-44.

Tye, M. (1995), *Ten Problems of Consciousness*, The MIT Press, Cambridge, MA.

Tye, M. (2000), *Color, Content, and Consciousness*, Cambridge, MA: MIT Press.

Van Gulick, R. (2001), "Inward and Upward: Reflection, Introspection, and Self-Awareness," *Philosophical Topics* 28: 275-305.

Van Gulick, R. (2004), "Higher-order global states (HOGS): An alternative higher-order model of consciousness," in R. J. Gennaro, ed., *Higher-Order Theories of Consciousness*, Amsterdam and Philadelphia: John Benjamins Publishers, 67–92.

Weisberg, J. (2003), "Being All That We Can Be: A Critical Review of Thomas Metzinger's *Being No One: The Self-Model Theory of Subjectivity*," *Journal of Consciousness Studies*, Vol. 10, No. 11, 89-96.