

Josh Weisberg

Being All That We Can Be

*A Critical Review of Thomas Metzinger's
Being No One: The Self-Model Theory of Subjectivity*

Some theorists approach the Gordian knot of consciousness by proclaiming its inherent tangle and mystery. Others draw out the sword of reduction and cut the knot to pieces. Philosopher Thomas Metzinger, in his important new book, *Being No One: The Self-Model Theory of Subjectivity*,¹ instead attempts to disentangle the knot one careful strand at a time. The result is an extensive and complex work containing almost 700 pages of philosophical analysis, phenomenological reflection, and scientific data. The text offers a sweeping and comprehensive tour through the entire landscape of consciousness studies, and it lays out Metzinger's rich and stimulating theory of the subjective mind. Metzinger's skilled integration of philosophy and neuroscience provides a valuable framework for interdisciplinary research on consciousness.

Metzinger's overall goal in *Being No One* is to defend a *representational* theory of subjectivity, one that reduces subjective mental processes to representational mental processes. Subjective experiences take place when there is a conscious perspective, an active first-person point of view. It occurs in an organism when 'there is something that it is like to *be* that organism — something it is like *for* the organism' (Nagel, 1974/1997, p. 519, emphasis in original). According to Metzinger, subjective experience emerges from the interaction of several kinds of conscious representational states: representations of the world, representations of the self, and representations that link the two.

Metzinger constructs his theory by first marking out what it is for a mental state to be consciousness, and what it is for an organism to be self-conscious. He holds that a mental state is conscious if it is globally available, it is integrated into

Correspondence:

Josh Weisberg, CUNY Graduate Center, Philosophy and Cognitive Science, 365 Fifth Avenue, New York, NY 10016-4309, USA. Email: jwsleep@aol.com

[1] Thomas Metzinger, *Being No One: The Self-Model Theory of Subjectivity*, Cambridge, MA: MIT Press, 2003, 699 pp., \$55, ISBN: 0262134179 (hbk). In this article, all page references are to this book unless otherwise stated.

an active representational world-model, and the organism cannot directly access the fact that the state is a representation. Self-consciousness occurs when conscious states are further integrated into a 'phenomenal self-model' that monitors, among other things, the various subject-object relations that the organism enters into. Mental states with these properties are experienced from a first-person perspective. They are subjective.

Metzinger employs what he calls 'the method of constraint satisfaction' to isolate a working concept of conscious experience. The method picks out constraints that are phenomenologically plausible but revisable identifying marks of mental state consciousness. Metzinger is sceptical of restrictive folk-psychological definitions of consciousness that focus exclusively on 'what it's like' to experience conscious mental states. He is also leery of approaches that rigidly operationalize the mind and therefore fail to take phenomenology seriously. Instead, his constraints are open to input from a range of interdisciplinary sources, yet reflective of the richness and complexity of conscious experience. In this way, Metzinger tries to avoid the pitfalls of simplistic reduction, while maintaining his distance from the mystery mongering of anti-physicalist claims.

Whether his approach is radically different from others is an open question. For example, Daniel Dennett's 'heterophenomenology' also allows for the input of phenomenological claims, with the goal of achieving a scientific explanation of consciousness (Dennett, 1991, chapter 4). So does Owen Flanagan's 'natural method' for studying consciousness (Flanagan, 1992). In addition, Metzinger's view of folk psychology is overly limited. Following the work of David Lewis (1972), folk psychology consists in a framework of causal platitudes that allow us to predict and explain one another's behaviour. This conception of folk psychology can provide a commonsense grounding for the theoretical constraints Metzinger employs. This, in turn, is useful in countering anti-physicalist intuitions that are allegedly rooted in common sense.

Nonetheless, Metzinger's approach is particularly noteworthy because of the attention he pays to empirical data. All of his constraints are explicated on a number of theoretical levels, with an eye towards their eventual incorporation into neuroscience. In addition, he runs aspects of his theory against various 'neurophenomenological case studies', culled from current research on the brain. This allows him both to refine his constraints in the light of empirical data, and to show the explanatory adequacy of his view. Indeed, the empirical data almost makes up a second book embedded in the first, and it provides a nice introduction to a number of arcane and interesting phenomena, from blindsight and agnosia to lucid dreams and out-of-body experiences. There is one important shortcoming with Metzinger's presentation of these empirical results, however. While the neurological data does offer some support to his view, and may refute 'classical theories of the mind' like Descartes' and Kant's, it would have been illuminating to consider rival *modern* theories of consciousness in this context. Rival modern views also claim to be compatible with the empirical data. Showing how the various theories come apart would be an important step towards deciding between them.

Metzinger delineates a number of constraints that pick out a working concept of mental state consciousness. I will focus only on those that are most central to his view. First and foremost among these is 'global availability'. Metzinger holds that a conscious state must be globally available for 'deliberately guided attention, cognitive reference, and control of action' (p. 118). Global availability is a third-person, functional claim about consciousness. Metzinger contends that there are two kinds of mental access involved in global availability, a nonconceptual kind of attention, and full-blown cognitive introspection. He argues for the presence of nonconceptual attention by citing empirical work on colour perception by philosopher Diana Raffman (1995). It is the nonconceptual form of attention that truly makes a mental state conscious (pp. 118ff, 615). When a mental state is suitably accessed, it can then guide context-sensitive, flexible action. In a number of places, Metzinger writes that global availability is what *makes* a state conscious (e.g., pp. 42, 117–18, 300, etc.), and that it is one of the few constraints that borders on a *necessary* condition for consciousness (p. 120).

There are substantive problems with the global availability constraint, however. First, it invokes a dispositional property, 'availability'. It is not clear that an available but unaccessed state is a conscious state. Metzinger himself notes that mental states 'attentionally *available* but . . . not attended *to* at all . . . [will] completely recede into the background of phenomenal experience' (p. 291, emphasis in original). This suggests that a state must be actively accessed to be conscious. And this certainly fits with the phenomenology: consciousness appears to be an active, occurrent phenomenon, from the first-person perspective.

But there is a more serious problem with global availability. There are many cases where states are available for attention, cognition, and control of action, but remain *unconscious*. Consider a musician who faces an impasse while composing a score. Sometimes, the solution will come as if from 'nowhere'. But clearly, the unconscious mind accessed the relevant representations to solve the problem. Here, there is presumably attentional access to sensory representations, and cognitive access to musical theory and aesthetic standards. But the representations were unconscious while the problem was being solved. Or consider cases of 'absent-minded' or distracted action, like walking while engrossed in conversation, or driving for long distances. These tasks certainly require directed attention and cognition, and clearly involve action. Again, many of the representations involved in the action remain unconscious. Finally, priming and subliminal perception can have complex effects on our behaviour. We must attend to and cognitively process the representations that generate the behaviour, yet they, too, remain unconscious.

Metzinger might argue that my description of these cases is driven by naïve folk psychology, and so should be discounted. But this folk-psychological intuition drives anti-reductionist claims as well. David Chalmers (1997) argues that no functional notion like availability can be identified with phenomenal experience. Likewise, Ned Block (1995; 2001) argues that global availability is nonphenomenal notion, and equating it with 'phenomenality' is a confusion.

These claims have bite because they trade on global availability's failure to fit with common sense.

Metzinger would be better advised to pick out access in terms of states that the subject is *aware of being in*. This claim fits with common sense: people generally do not consider a state conscious if they are totally unaware of being in it. And the claim has received theoretical treatment and defence from 'higher-order representational' theorists like David Rosenthal (1986; 1997; 2002a) and William Lycan (1996; 2001). We can recast the availability constraint as the claim that conscious states are actively accessed mental states that the subject is aware of. This is offered as a positive refinement of Metzinger's constraint, rather than a knock-down objection. It allows for Metzinger's theoretical account to proceed largely as before, but it fits with common sense in a way that helps fend off anti-physicalist intuitions.

Another important point relating to the global availability constraint concerns the *kind* of access that we have to our conscious states. Metzinger argues that it is primarily a *nonconceptual* form of attention. According to Metzinger, without this kind of access, mental states are not conscious (p. 615). His argument for this point relies on Diana Raffman's work (1995). Raffman shows that we can discriminate many more colours from one another than we can recognize and reidentify individually. Concept possession arguably entails the ability to reidentify a category over time. Yet we can be conscious of discriminatory differences that we can't reidentify. Therefore, we must possess an awareness of the colours that is nonconceptual. On this basis, Metzinger posits a form of nonconceptual attention that allows us to discriminate between various sensory qualities (pp. 118ff).

But this does not follow. Certainly, there must be a level of sensory representation that picks up the subtle differences in colour that we can perceive (for example, the 'sensory qualities' posited by Austen Clark, 1993; 2000). But our *access* to these representations can still be conceptual. We possess concepts of the *similarities and differences between sensory qualities*, like 'lighter than' or 'darker than'. When accessing subtle differences between represented stimuli, we can *conceptually* be aware of them in terms of their similarities and differences (see Rosenthal, 1999; 2002b). We need not, and indeed do not, possess individual concepts for the qualities, as Raffman establishes. But this does not entail that we do not *access* the qualities conceptually. In addition, there are theoretical difficulties in spelling out a theory of nonconceptual content (see, e.g., Stalnaker, 1998). Raffman's work does not rule out conceptual awareness of our mental states, and it may be theoretically advantageous to posit wholly conceptual access.

Metzinger's next constraint holds that conscious states are always embedded in a 'world-model', a coherent, unified 'representation of reality as a whole' (p. 131). Conscious content is always experienced as 'being in a world', as embedded in a unified structure in which the conscious organism is *situated*. Metzinger calls this the 'globality' constraint, and he attempts to link this phenomenal constraint to the functional global availability constraint. He claims that

to be part of a world-model is to be accessible by a number of different systems (see, e.g., pp. 138, 275). But this connection is not obvious. To be integrated into a phenomenal 'world' does not entail that a state is globally available for higher-order processing. Indeed, the 'buzzing, blooming confusion' of our phenomenal world might make states less accessible. Furthermore, what of punctuate experiences like a flash of sensory memory, or a random, drifting, conscious thought? It is not clear that these states are really integrated in the manner that Metzinger describes. All conscious states are arguably unified in that they are experienced by one subject, but this is a matter of how they are accessed by the self, not of how they are integrated into a world-model.

The next constraint is vital to Metzinger's theory. It is called the 'transparency constraint', and it helps clarify a number of puzzling features of consciousness. A representation is transparent, in Metzinger's sense, when the 'representation cannot be recognized *as* a representation by the system itself' (p. 131). The content of such a representation appears to the subject as real, as 'given'. This is because, as subjects, we are not aware that we are the authors, as well as the readers, of our conscious representational contents. This creates a situation that Metzinger terms 'autoepistemic closure'. We cannot directly access that we are engaged in a representational process, and that we experience the world *through* representational states. Such information is 'closed' to us.

Transparency explains the apparent immediacy of conscious states. They seem directly 'given' because the processing stages that led to their presence are 'attentionally unavailable' (p. 165). But this direct awareness is only apparent. A variety of representational processes underwrite conscious experience, but they remain unaccessed. This results in the appearance of immediacy. This 'special kind of darkness' also helps explain certain anti-physicalist intuitions that pervade the study of consciousness (pp. 330f). We can't directly access the fact that we are representational systems. Therefore, we find completely unintuitive any theory that makes such a claim.

Some conscious states are 'opaque' (as opposed to transparent), according to Metzinger. For example, propositional attitudes and types of mental simulations are not transparent because we know that we are the initiators of these states (p. 174). But it is not clear that these states are actually opaque. Are we ever really aware of 'earlier processing stages' that lead to the tokening of propositional attitudes? Rather, these conscious states just appear to us like other conscious states, except that they are accompanied by the (usually tacit) belief that we are the initiators of the states. But this falls short of losing transparency, which involves awareness of processing and construction (pp. 165f). In any event, transparency, and the autoepistemic closure it entails, explains how we could be representational systems, and yet remain unaware of it. Our conscious states are just given to us. The fact that they are representations is not. It is this gap in appearances that the anti-physicalist attempts to widen into a metaphysical chasm.

As of yet, conscious mental states are not *for* anyone; they are not subjective. This requires the satisfaction of the 'perspectival constraint', which involves the

creation and activation of a 'phenomenal self-model' (PSM). A PSM is a complex representational state that makes 'system-related information' available for the organism (p. 300). According to Metzinger,

The content of the PSM is the content of the conscious self: your current bodily sensations, your present emotional situation, plus all the contents of your phenomenally experienced cognitive processing. . . . All those properties of yourself, to which you can now direct your attention, form the content of your current PSM (p. 299).

We experience ourselves *through* this representational model. It provides us with our sense of self, our sense that we 'own' our mental lives. Its existence allows for the felt experience of what Metzinger calls 'mineness.' This phenomenal property emerges when conscious contents are integrated into the PSM. Contents thus integrated into the PSM are experienced as part of us (pp. 302, 306ff).

The crucial step in the transition from self-model to phenomenal self-model is the satisfaction of the transparency constraint (pp. 330–7). When a transparent self-model is available for us, 'we do not experience the contents of our self-consciousness as the contents of a representational process . . . but simply as *ourselves, living in the world right now*' (p. 331, emphasis in original). Metzinger claims, 'If all other necessary and sufficient constraints for the emergence of phenomenal experience are satisfied by a given representational system, the addition of a transparent self-model will by necessity lead to the emergence of a phenomenal self' (p. 337).

When we possess an active PSM, we can become aware of ourselves *as* ourselves, as individuals and as active agents. We can also differentiate ourselves from the world, and from other agents. The presence of a self/other (or subject/object) boundary requires a self-model, and the PSM makes the boundary available for cognition and action. It also makes it available for conscious representation of the various subject–object relations we enter into. This special kind of representation is the 'phenomenal model of the intentionality relation' (PMIR), and it is at the heart of Metzinger's theory of subjectivity.

According to Metzinger, in order to arrive at 'full-blown subjective consciousness,' three steps must be taken. First, we must generate a world-model, then a self-model, and finally, we must achieve 'the transient integration of certain aspects of the world-model *with* the self-model' (p. 427). The PMIR achieves this last step. It is a representation of the self mentally interacting with the world. Some examples of the content of a typical PMIR are, 'I am someone who is visually attending to the color of the book in my hands', or 'I am someone currently grasping the content of the sentence I am reading', or 'I am someone now deciding to get up and get some more juice' (p. 411). The PMIR 'depicts a certain *relationship* as currently holding between the system, as transparently represented to itself, and an object component' (p. 411, emphasis in original). When this occurs, a perspective is created and subjective experience results. The PMIR also represents the 'directedness' of perception and cognition. It creates an awareness of what Metzinger calls 'the arrow of intentionality', of our asymmetrical representational relation with the world (p. 416). The PMIR can even loop back towards

the self-model and represent it as an object. This creates 'full-blown' self-consciousness.

Metzinger holds that all subjective experience involves a PMIR (p. 427). However, this is problematic, because it entails that whenever we are conscious, we are conscious *of* the PMIR. It is *phenomenal*, after all. But this does not seem right. Right now I am staring (glassy-eyed) at my computer screen. I am conscious of the screen, and conscious of what I'm typing. But in addition, Metzinger holds that I'm conscious of 'myself in the act of attending to the screen', and 'myself in the act of understanding my words', etc. But I am rarely aware of these things, unless I actively engage in introspection. Then I will be conscious of them, but until then they are not present in consciousness. If they were, consciousness would be much more cluttered than it is. Ordinarily, all we consciously experience is the world and the things in it.

This is not to say that Metzinger isn't correct about the role that the PMIR plays in subjective experience. He is simply incorrect about the degree to which we are ordinarily conscious *of* the PMIR itself. It makes theoretical sense, therefore, to posit a *nonconscious* model of the intentionality relation (NMIR). There is no reason to think a NMIR cannot achieve the various theoretical tasks that Metzinger enumerates for the PMIR, and it possesses the advantage of correctly fitting with our phenomenology. Metzinger may argue that he is in fact often aware of the PMIR. But that fits with my proposal. The NMIR is nonconscious, but poised to be accessed by introspection, and thus made conscious. The NMIR is available, but nonconscious.

We are now in a position to state Metzinger's self-model theory of subjectivity. He holds that '*Phenomenally subjective* experience consists in transparently modeling the intentionality relation within a global, coherent model of the world' (p. 427, emphasis in original). When phenomenally subjective experience is instantiated, a first-person perspective exists. The theory is fully representationalist in that it employs only representational 'models' and their properties in its formulation. Please note that Metzinger elucidates several additional constraints, including ones that deal with the temporal and qualitative features of conscious mental states. He also provides an extensive discussion of mental simulation and the evolutionary function of consciousness. It is beyond the scope of this review to address all of these issues, though I have done my best to touch on the most important constraints.

In an interesting twist, Metzinger endorses an *eliminativist* stance towards a number of mental structures, including *the self*. He argues that our folk concept of the self picks out an individual substance that persists through time and underwrites our identity. However, all we find in the mind are dynamic, shifting psychological states. There is nothing answering to this commonsense picture of the self. Because we transparently represent the self, we are 'autoepistemically closed' to the fact that we only have direct access to our *representational model* of the self. It represents a range of invariant features of our body, emotional state, and memory, and we interpret this, due to transparency, as a substantial existent. We 'reify' the self as a substantial object. But this leap is unwarranted, and

modern science, in unity with Metzinger's theory, will have no place for a substantial notion of self. Thus the title of his book is *Being No One*.

But this conclusion is too strong. There are many cases where we have revised a concept, even a folk concept, over time. Though the ancients may have thought that the stars were holes in the sky, this does not mean that there are no stars, or that they were referring to something different from us as they gazed at the heavens (see Rosenthal, 1980; Stich, 1996). Why not say that the self is just those invariant elements of the mind that the PSM picks out? Various characteristics of our bodies, moods, emotions and memories remain stable over time, and these underwrite our personal identity. These elements can be construed as the self, even if they do not answer to all our folk intuitions. This proposal may do damage to the title of the book, but not, I believe, to the spirit of the theory.²

References

- Block, N. (1995), 'On a confusion about a function of consciousness,' *Behavioral and Brain Sciences*, **18** (2), pp. 227–47.
- Block, N. (2001), 'Paradox and cross purpose in recent work on consciousness,' *Cognition*, **79** (1–2).
- Block, N., Flanagan, O. and Güzeldere, G. (1997), *The Nature of Consciousness: Philosophical Debates* (Cambridge, MA: MIT Press).
- Chalmers, D.J. (1997), 'Availability: The cognitive basis of experience?' in Block *et al.* (1997).
- Clark, A. (1993), *Sensory Qualities* (Oxford: Clarendon Press).
- Clark, A. (2000), *A Theory of Sentience* (Oxford: Oxford University Press).
- Dennett, D.C. (1991), *Consciousness Explained* (Boston, MA: Little Brown).
- Flanagan, O. (1992), *Consciousness Reconsidered* (Cambridge, MA: MIT Press).
- Lewis, D.K. (1972), 'Psychophysical and theoretical identifications,' *Australasian Journal of Philosophy*, **L**(3), pp. 249–58.
- Lycan, W.G. (1996), *Consciousness and Experience* (Cambridge, MA: MIT Press/Bradford Books).
- Lycan, W.G. (2001), 'A simple argument for the higher-order representation theory of consciousness,' *Analysis*, **61** (1), pp. 3–4.
- Nagel, T. (1974/1997), 'What is it like to be a bat', reprinted in Block *et al.* (1997).
- Raffman, D. (1995), 'On the persistence of phenomenology' in *Conscious Experience*, ed. T. Metzinger (Exeter: Imprint Academic).
- Rosenthal, D.M. (1980), 'Keeping matter in mind,' *Midwest Studies in Philosophy*, **V**, pp. 295–322.
- Rosenthal, D.M. (1986), 'Two concepts of consciousness,' *Philosophical Studies*, **49**, pp. 329–59.
- Rosenthal, D.M. (1997), 'A theory of consciousness,' in Block *et al.* (1997).
- Rosenthal, D.M. (1999), 'Sensory quality and the relocation story,' *Philosophical Topics*, **26** (1–2), pp. 321–50.
- Rosenthal, D.M. (2002a), 'Explaining consciousness', in *Philosophy of Mind: Contemporary Readings*, ed. D.J. Chalmers (New York: Oxford University Press).
- Rosenthal, D.M. (2002b), 'Consciousness and higher-order thought', *Encyclopedia of Cognitive Science* (Hampshire: Macmillan Publishers, Ltd.).
- Stich, S.P. (1996), *Deconstructing the Mind* (New York: Oxford University Press).
- Stalnaker, R. (1998), 'What might nonconceptual content be?', in *Concepts, Philosophical Issues*, Vol. 9, ed. Enrique Villanueva (Atascadero: Ridgeview).

[2] Thanks to David Rosenthal for comments on an earlier draft.