PSYCHE

Reply to Weisberg: No direction home—searching for neutral ground

Thomas Metzinger
Philosophisches Seminar
Johannes Gutenberg-Universität Mainz
D-55099 Mainz
www.philosophie.uni-mainz.de/metzinger
metzinger@uni-mainz.de

PSYCHE 12 (4), August 2006

Reply to: Weisberg, J. 2005. Consciousness Constrained: A Commentary on *Being No One*, *Psyche* 11 (5).

Keywords: Method of interdisciplinary constraint satisfaction (MICS), global availability, global workspace theory (GWT), perspectivalness, phenomenal model of the intentionality relation (PMIR)

I have learned a lot from Josh Weisberg's substantial criticism in his well-crafted and systematic commentary (see also his book review in Weisberg 2003). Unfortunately, I have to concede many of the points he intelligently makes. But I am also flattered by the way he ultimately uses his criticism to emphasize some of those aspects of the theory that can perhaps possibly count as exactly the core of my own genuine contribution to the problem—and nicely turns them back against myself. And I am certainly grateful for a whole range of helpful clarifications.

First, Weisberg has finally given a name to what I am actually *doing*, to the approach I have developed: MICS ("Method of Interdisciplinary Constraint Satisfaction"). I am relieved that he did not directly attack what I myself see as possibly the greatest weakness of my own approach. Is the top level of description—the employment of first-person phenomenological constraints—really a "discipline," particularly for a philosopher who claims that, strictly speaking, no such things as "first-person data" exist? In what sense is this really inter*disciplinary* constraint satisfaction? I now have a name for my own approach, but am also immediately confronted with the main danger—namely of "operationalizing away the difficulties" (p. 3), of being overly impressed by specific empirical models of consciousness, and then of importing the implicit theoretical assumptions of these models. By drawing these assumptions all the

way up into the phenomenological level of description, I run the risk of contaminating it by suddenly discovering phenomenological features which, strictly speaking, never really belonged to the original common-sense description of the target phenomenon. Weisberg has an excellent point here. Just like Allan Hobson, he criticizes me for being a bit *too* ecumenical. Not only do I draw on empirical theories, which are in themselves controversial (as Hobson argues), but by endorsing a particular classical model on the market, namely the Baarsian global workspace theory, I import theoretical background assumptions from the functional level of description and then dubiously "rediscover" them in my own phenomenology. Here is Josh Weisberg's diagnosis: "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 (…) 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" (p. 5).

True. But wasn't Paul Churchland also right in his prediction that applying a new neuroscientific terminology to our own inner states in introspection would actually *enrich* conscious experience itself? True, MICS-style neurophenomenology runs the risk of importing "substantial theoretical claims directly into the data" (p. 6), but first, there may be no "data" in a stricter sense. Second, does the beauty of the neurophenomenological approach not also consist in the fact that it *changes* our own phenomenal experience as we proceed?

Take as an example my own criticism of the much too broad and undifferentiated notion of "global availability." On page 9 of his commentary, Weisberg quotes my own attempt to differentiate the concept in terms of availability for introspective attention, cognitive reference, and motor selection (BNO: 31). Developing these conceptual constraints has certainly changed my own introspective experience. Not only was I inspired by Diana Raffman (1995) and Daniel Dennett (1988), read up on some perceptual psychology, and uncritically imported low-level theoretical assumptions into the top level of description. This also changed my own phenomenology *itself*: I am now much more acutely aware of the fact that there are subtle, ineffable nuances in my own sensory perception of the world, and of my own inability to form mental predicates for, say, the myriad maximally determinate shades of color that make up my phenomenal world. I have discovered something new. Something that arguably does *not* belong to the common-sense theory of consciousness Weisberg advocates as a neutral ground for defining the explanandum. Most people are surprised when their attention is drawn to the fact that they do *not* have qualia in the classical sense introduced by C.I. Lewis, that they can*not* identify their most simple sensory contents in perceptual experience across time. Isn't this a good argument against the possibility that everyday folk phenomenology, including its common-sense descriptive systems, could really provide a "neutral ground" for the scientific investigation of consciousness?

Weisberg is certainly right that at an initial stage of a complex research program, it is good to first establish some common ground in terms of a pretheoretical characterization of our epistemic targets. On page 6, he writes, "if a theory cannot explain why the explanandum appears as it does to common sense—that is if it cannot "save the common-sense appearances"—we do not count it as a successful theory." My prediction is that our future theory of consciousness will completely destroy these "common-sense

appearances," although it may also explain where our theoretical intuitions actually come from. Again, Weisberg is probably right when he writes that "to begin, we must have a handle on the phenomenon as we ordinarily pick it out—otherwise we can't be sure that we have explained the features that interested us at the outset of theorizing" (p. 7). However, we are not at the outset anymore—the investigation of consciousness has *already* started on a global, industrial scale. The difficulty is that, at least for the experts, the target phenomenon itself has already begun to change. I do not want to discuss the issue of whether the common-sense taxonomy ever presented us with a neutral starting ground to frame the problem in the first place. But even if this had been the case in the past—experts like the participants in this symposium would have departed from this neutral ground long ago. We have all thought long and hard about the problem of consciousness and developed our own conceptual systems, and in doing so we have more or less subtly changed the phenomenological landscape of our own inner experience as well. We are already beyond the initial stage, and from Weisberg's perspective, we find ourselves lost in the middle of a terrible swamp. This is progress.

Walking through the streets of Brooklyn while contemplating philosophical progress (p. 10) certainly involves the possession of active representational contents that are available for behavioral control. What they, and the philosopher harboring them, lack is the extra flexibility associated with maximal *context-sensitivity*. This is exactly what global workspace theory would predict from the functional level of description and from the everyday phenomenology of a philosopher deeply immersed in an *internal* context. We can certainly support this claim. When Weisberg says that unconscious percepts must be available to "control" his actions in order to ensure that he arrives without injury, however, he slips into a subtle mereological fallacy: he ascribes agency (personal-level control) to subpersonal states. But subpersonally controlled body movements could hardly count as "actions". He is correct in claiming that the presented case has a clear folk-psychological reading in terms of nonconscious states, and in asking, "What reason do we have to deny these interpretations?" (p. 10) We have philosophical reasons: they are not *coherent*. Unconscious behavioral control is a process of dynamical self-organization. The relevant control parameters are available, but not currently accessed. How could one, as Weisberg demands, effectively delineate those types of behavior that exhibit the requisite flexibility? By defining them in terms of the reaction time to novel stimuli, for example. This is an empirical question, not a philosophical one. Or is it?

Weisberg continues by making a number of further interesting observations. For instance, he claims that access to introspections is not constitutive of consciousness, but rather that there is something about conscious states that renders them accessible to introspection (p. 11). A typical folk-phenomenological idiom corresponding to this description would be that we can "make them conscious." In terms of cognitive ergonomics, this manner of speaking may be perfectly fine. But please note how, on closer inspection, this notion is itself incoherent: As an agent, you can only operate on something that already belongs to your model of reality, to your world. A strictly unconscious piece of information could never be something we as conscious subjects could actively "make conscious"—we wouldn't know how to look for it. Weisberg also points out how minimally conscious states (which would have to satisfy the presentationality constraint as well) would not be relevant *for* a subject and therefore intuitively would not be conscious. This is an important observation, a point that has led

to a lot of confusion in scientific debates on consciousness. True, minimal consciousness without perspectivalness is not part of our everyday notion of consciousness, and in *this* sense, the conjunction of the three first constraints would not be sufficient for consciousness. But does this really imply that there is no fact of the matter? Another interesting observation Weisberg makes is that there are types of phenomenal content—the pain caused by an anvil dropping on your foot—whose causal role seems to be associated with an important evolutionary function. Let us call this function "attention fixation": pain and strong emotions interestingly lead to a functional *rigidification* by blocking the focus of attention and forcing of the subject to do something about the cause of his pain. Here, I concede to Weisberg that these states satisfy the phenomenological globality constraint (they are part of my own world), and that with regard to these states, I do *not* subjectively experience what I have termed "my own selectivity and flexibility in dealing with them."

Next, Weisberg asks another interesting question about the global phenomenal property of "being in a world": Wouldn't integration actually reduce the salience of stimuli, by blending them in with the "buzzing, blooming confusion" of an experienced world (p. 11)? This is a good question. And one that is interestingly mirrored on the level of microfunctional analysis, namely in searching for a dynamic integration function for the unity of consciousness. For instance, a global synchronization process could easily cause what network theorists call a "superposition catastrophe." By wiping out all differentiation, it could lead to a global state that would not satisfy what, in BNO, I termed the "convolved holism" constraint (a dynamic, flexible hierarchy of nested phenomenal wholes), but that instead would resemble an epileptic seizure. As we see, Weisberg has a number of good arguments for his claim that the functional reading of global *availability* is not a necessary ingredient for a pretheoretical characterization of consciousness. But the phenomenological-level globality constraint as such (and it was my mistake to have made it appear as an independent constraint, which it was not meant to be; see Weisberg p. 11; BNO: 131-143) certainly fulfils this condition. Isn't it true that, not only in a pretheoretical sense, the essence of consciousness is exactly the *appearance of a world*?

Josh Weisberg obviously is more of a Kantian than I am, as we see in section 5 of his commentary. For him, the essence of our target phenomenon seems to be the emergence of a *subject*, or a transparent model of the intentionality relationship, as I have called it (see also Metzinger 2006 for a recent application of the concept). I must admit that I find this strategy of changing the "relevance landscape" in the original set of constraints in order to critically turn it against my own position extremely interesting. Weisberg writes, "I contend that taken together, transparency and perspectivalness form a well justified working concept of consciousness" (p. 12). However, this concept would be circular: transparency is a property of phenomenal states only, and an unconscious state is neither transparent nor opaque. Consequently, we would import phenomenality into our working concept at the very beginning.

It is possible and interesting to investigate unconscious versions of the internal model of the intentionality relationship. If we want to satisfy the "acquisition constraint", that is, if we want to understand how this high-level mental structure could *gradually* come into existence in the course of natural evolution or in childhood development, then

it makes great sense to ask: Do simpler, perhaps unconscious, precursors exist in the brain? Together with Vittorio Gallese, I have offered some empirical ideas about the unconscious, evolutionary precursors of the PMIR (see Metzinger and Gallese 2003), of the nonphenomenal modeling of organism-object relations. And in footnote xxi, Weisberg himself discusses the concept of a nonconscious model of the intentionality relation (NMIR), and in the spirit of David Rosenthal writes that "the highest-level MIR must be nonconscious to block the regress." It is certainly a highly original idea to substitute the global integrational function in my own HOB (highest-order binding)-model (Metzinger 1995) with the idea of a highest-level MIR constituting the unity of consciousness. But in the same footnote, Weisberg writes, "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." Unfortunately, given the conceptual framework developed in BNO, this statement contains two contradictions: first, a PSM doesn't *become* conscious, it *is* conscious; second, an unconscious structure like an NMIR could not represent it in a "transparent way," because this is something only phenomenal states can do. And this observation points to the central difficulty: NMIRs may exist and help us analyze the concept of phenomenal perspectivalness, but transparency as such would make our working concept of consciousness circular. The connection Weisberg makes between Rosenthal, Lormand, the transitivity principle, and Thomas Nagel (1974: 519) is lucidly observed and goes to the heart of the matter: the PMIR, or so I propose, is exactly what could lead to a functional analysis of what a first-person perspective is in the sense introduced by Nagel.

I disagree that we could be aware of a specific belief without introspectively accessing the processuality leading to its activation: isn't it true that we would then slide into a manifest daydream, into a fully absorbed process of conscious cognition that would not be experienced *as* cognition anymore? Transparency is a necessary feature for the phenomenal property of *cognitive agency*, for the experience of being a thinking self, actively forming and selecting cognitive contents. It is the transparency of a residual self-model, which is a necessary feature of perspectival consciousness, that Weisberg is looking for. I fully agree however, that one of the more important desiderata for the future is a theory that distinguishes different kinds of phenomenal opacity, in different domains and relative to different access mechanisms. There may be more than one kind of opacity with regard to the phenomenology of conscious thought (e.g., attending to the process of forming and disambiguating contents over time vs. having occurrent metarepresentational beliefs *that* one is currently thinking), and the sensory opacity involved in my initial example of a visual pseudo-hallucination (breathing, abstract geometrical patterns on the wall in front of you) may turn out to be a totally different phenomenon. For a number of reasons, I will not enter into a discussion about the phenomenology of selfless religious experience in this short reply. Instead, let me note one last point of disagreement with Weisberg.

Making the transparent PMIR the centerpiece of a neurophenomenological theory of consciousness, and this is Weisberg's interesting proposal, would dramatically shift our view of the phenomenology. For instance, theories about the "fringe," i.e., about everything that is not a component of self-consciousness or the focus of experience, would now become much more important. But I think Weisberg may underestimate how dramatic a turn in our general idea of consciousness this would actually be. It is simply

not true that you can get "being in a world," the phenomenology of globality, from the representational contents of the PMIR. All we get is a comprehensive representation of the system as standing in relation to *parts* of this world. We never have an "upward mereology." That is to say, we cannot describe the phenomenal subject as being *embedded* in a global whole anymore, only as currently being *directed* at an object component. There would no longer be a world-system relationship in the conscious model of reality. If we followed Josh Weisberg's interesting line of argument, we would lose the globality of subjectivity. We would give up the pretheoretical intuition of "being in a world" as the essence of conscious experience in favor of another strategy to fix the data: we would now be searching for consciousness as "being a dynamically directed self."

References

Dennett, D.C. (1988). Quining qualia. Marcel, A., and Bisiach, E., eds. (1988). *Consciousness in Contemporary Science*. Oxford: Oxford University Press.

Metzinger, T (1995). Faster than thought. Holism, homogeneity and temporal coding. In T. Metzinger (ed), *Conscious Experience*. Thorverton: Imprint Academic & Paderborn: mentis.

Metzinger, T. (2006). Conscious volition and mental representation: Towards a more fine-grained analysis. In Sebanz, N., and Prinz, W., eds., *Disorders of Volition.* Cambridge, MA: MIT Press.

Metzinger, T., and Gallese, V. (2003). Motor ontology: the representational reality of goals, actions, and selves. *Philosophical Psychology* 16: 365-388.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review* 83: 435-50.

Raffman, D. (1995). On the persistence of phenomenology. In Metzinger, T, ed., *Conscious Experience.* Thorverton, U.K.: Imprint Academic.

Rosenthal, D. (2002). *Consciousness and Mind*, Oxford: Clarendon Press.

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* 10 (11), p. 89-96.