



Metzinger's Matrix: Living the Virtual Life with a Real Body

Shaun Gallagher
Department of Philosophy and
the Cognitive Sciences Program
University of Central Florida
Orlando, FL 32816-1352
U.S.A.
© S. Gallagher
gallaghr@mail.ucf.edu

PSYCHE 11 (5), June 2005

Keywords: embodiment, disembodiment, self, Cotard's Syndrome

Commentary on: Metzinger, T. 2004. *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.

ABSTRACT: Is it possible to say that there is no real self if we take a non-Cartesian view of the body? Is it possible to say that an organism can engage in pragmatic action and intersubjective interaction and that the self generated in such activity is not real? This depends on how we define the concept "real". By taking a close look at embodied action, and at Metzinger's concept of embodiment, I want to argue that, on a non-Cartesian concept of reality, the self should be considered something real, and not simply an illusion.

1. Introduction

Metzinger's rich and highly detailed book overflows with metaphors that stretch our imagination for purposes of thinking about the self. Metaphors have, he suggests, "great heuristic fertility" (Metzinger 2004: p. 51). It is clarifying to think of ourselves metaphorically as neurophenomenological cavemen who live in empty Platonic caves—where 'empty' means *we* are not even there (pp. 547ff). Alternatively, we are, metaphorically, little red arrows on subway maps, although, phenomenologically, we are neither red (we are transparent in Metzinger's special sense) nor arrows (we only function like arrows of intentionality), and the subway map is actually an internal subpersonal map (pp. 551ff). Likewise, he suggests, we are like total flight simulators, generating our

virtual world in real-time, and with such sophistication that we trick ourselves into thinking we are living in a real world (p. 553). Metzinger does not shy away from mixing these metaphors, in an attempt to gain even more heuristic clarity. He writes: "Like a seriously deluded tourist who actually believes he is the little red arrow, the caveman is like an airplane that functionally owns its hardware, but has only just begun to appropriate the *simulator*. Neurophenomenologically, he is a shadow boxer who has become hypnotized by his own internal shadow" (p. 557).

In Metzinger's book one frequently finds a subtle slippage between metaphor and science. For example, he proposes to employ a "new metaphor," namely that "phenomenal experience during the waking state is an *online* hallucination" (p. 51). Metzinger turns this metaphor immediately into scientific fact: "As a matter of fact, some of the best work in neuroscience ... suggests a view of the human brain as a system that *constantly* simulates possible realities In our present context, a fruitful way of looking at the human brain, therefore, is as a system which, even in ordinary waking states, constantly hallucinates at the world ... thereby generating the content of phenomenal experience" (pp. 51-52). Setting aside the possibilities of an evil demon or an advanced race of artificial agents, either Metzinger is right and we are all caught in a matrix of virtual life that is "self" (= brain)-generated, and "no such things as selves exist in the world" (p. 1), *or* Metzinger has become tangled in a matrix of his own self-generated metaphors. As in all the modern matrix scenarios, the question is whether there is a way to test out such conclusions, especially when it comes to our own existence. For if Metzinger is right, then, as he suggests, even Descartes's surety about his own existence is open to doubt, since in some possible world there could be a Descartes suffering from Cotard's Syndrome who intuits that he does *not* exist (e.g., p. 456). Apparently, and reasonably, and likely by evolutionary design, most of us are tempted to take the word of the real Descartes over the word of the possible (and possibly inexistent) Descartes. Furthermore, I think that most of us are tempted to ask whether being no one in a virtual or hallucinated world is really (or at least virtually) very different from being a philosophical Cotardian.

For Metzinger, science is the measure of reality. If science demonstrates (although in this regard it is not clear that science does demonstrate rather than simply assumes) that certain things (the physical world, or brains, for example) are real, then they are. On this conception of reality, one would think that one's own body is something real. And according to some theories of personal identity, that body is *what* I am, even if it is not *who* I am. My existence and my body seem to be inextricably linked; isn't this something that makes me real? For Metzinger, things are much more complex. Consider two of his claims, made within the space of a few pages, in a very "pagey" book.

The conscious self-model in the caveman's brain itself is a phenomenal self characterized not only by full-blown prereflective embodiment but by the comprehensive, all-encompassing subjective experience of *being situated*. (p. 552).

[A] disembodied but appropriately stimulated brain in a vat could – phenomenologically – enjoy exactly the same kind of conscious experience as you do right now while reading this book. In principle, it would even suffice to properly activate just a subset of this brain, the minimally

sufficient neural correlate of your present state, to make a 'phenomenological snapshot' of exactly the same kind of conscious experience emerge" (p. 547; also see pp. 295, 335, 462).

"Full-blown prereflective embodiment" is an impressive phrase, but in this context it seems to count for little if it doesn't add anything to our brainy experience. I want to look closely at Metzinger's discussion of embodiment, and disembodiment, and specifically ask: What does "full-blown prereflective embodiment" mean, and what role does it play in our experience?

2. Feeling disembodied

For Metzinger, as for science, animal bodies are definitely things that exist, with the status of the real. For purposes of this paper, I'll define 'real' as anything that has the same ontological status bestowed by naturalism on things like brains, biosystems, and scientific equipment. For Metzinger, I believe, bodies are real in this sense since throughout his book he refers to important biological processes in the human body. Thus, he claims that the phenomenal self-model "is firmly anchored in the autonomous bodily dynamics of elementary bioregulation" (p. 505; see pp. 285ff., 485). These dynamics, he notes, depend on "complex homeodynamic self-regulation in the upper brainstem and hypothalamus, and a large number of different receptors external to the brain but internal to our body, which are, for instance, found in skin, muscles, joints, and viscera" (p. 439). Metzinger accepts the distinction between body image and body schema (as I have argued for it, see Metzinger, p. 485ff), equating the body image with the *phenomenal* self-model (PSM) and the body schema with the *functional* self-model. He also adopts a phenomenological distinction between *Körper* and *Leib*. One way to frame our question is this: Is real, full-blown pre-reflective embodiment equivalent to objective embodiment (*Körper*) or lived embodiment (*Leib*), or some combination of these?

Now a reader who is familiar with the work of Edmund Husserl needs to be careful here, because they may be inclined to think of this distinction as the one made by Husserl (1952). Metzinger, however, refers to this distinction as a "traditional conceptual distinction" available in the German language (p. 457). And he explains it as follows: "A *Leib* is a lived body, one that is connected to a soul, or, in more modern parlance, the body *as subject*, as the locus of an individual first-person perspective" (p. 457). This is consistent with Husserl's phenomenological concept of *Leib*. Metzinger characterizes the *Körper*, however, as an "inanimate object," and specifically as the dead corpse, which the Cotard patient experiences himself to be. In contrast, for Husserl, the *Körper* is not a dead body, but the Cartesian objective body as it is studied by science, or as it is experienced by someone who, for various reasons, may experience their own body as an objective thing, "from the outside" so to speak. When Descartes talks about the body, he tends to talk about it as an objective body in this sense. As such, however, the *Körper* is animate and alive, the objective version of the very biological body that is also the body *as subject* – something that Husserl clearly sees, but Descartes, who, on his own account, has only unclear ideas about this, does not.

We can appeal again to the distinction between the Cartesian thinker, who philosophically understands the body to be the objective body, and the Cotardian who thinks the body is dead. Both of them seem to overlook the lived body. Does this mean

that Cotard's Syndrome falls into the category of feeling disembodied, as Metzinger suggests (p. 457)? He points out that there are a number of phenomenological states in which someone may experience themselves as disembodied. This would include cases of deafferentation, such as Oliver Sack's "disembodied" woman, Christina, and Jonathan Cole's patient, IW (Metzinger pp. 439-43, 457; see Gallagher and Cole 1995). Cole, however, suggests that rather than disembodied, such patients are better described as super-embodied, since they are much more conscious of their bodies than are normal subjects, i.e., they have an enhanced body image. But clearly, in some sense Cotard's Syndrome may fall into the category of feeling disembodied. For the Cotard patient, this means that he has "a bodily self-model as a *Körper*, but not as a *Leib*" (p. 457). So, in this case, to be disembodied is to lose the sense of the lived body.

Let's consider several different ways we could speak of disembodiment.

- (1) *Cartesian* (philosophical) disembodiment: Descartes thinks of his body as an objective body, a *res extensa* (the Cartesian *Körper*), and fails to recognize his body as a lived body, the body as subject.
- (2) *Cotardian* disembodiment: the Cotard patient thinks of his body as a dead object (the Cotardian *Körper*), and fails to recognize his body as a lived body.
- (3) *Out-of-body* disembodiment: on the one hand the subject experiences her body from the outside, as either a Cotardian *Körper* or a Cartesian *Körper* – either dead or alive. In this sense her experience is disembodied in either sense (1) or sense (2). On the other hand, in regard to the lived body, the subject either (a) experiences herself as an embodied *subject* (as having a perceptual perspective floating above the scene, for example), or (b) experiences herself as "being in no body," and in this sense feels completely disembodied. As Metzinger puts it, a subject in this latter case is either "weakly embodied" or has no "explicit body image" (p. 499).
- (4) *Brain-in-a-vat* disembodiment: in this situation it seems that there would be *real* disembodiment, but, oddly, the subject would have a quite normal embodied experience—she would not experience her body from the outside in any abnormal way, and she apparently would have an explicit body image. She would "enjoy exactly the same kind of conscious experience as you do right now," assuming you, the reader, are not Cotardian or Cartesian, or having an out-of-body experience.

On Metzinger's view, having no explicit body image seems to be equivalent to feeling disembodied in sense (3b). The absence of an explicit body image, however, does not necessarily involve feeling disembodied. In (3b), the person experiencing out-of-body disembodiment explicitly feels the lack of her body—she has a positive experience of *not* being in her body. She *feels* disembodied. In contrast, a patient suffering from unilateral neglect does not have an explicit body image for the neglected side, but does not experience this as a lack, or as *not* being in that side of her body. He does not *feel*

disembodied. Moreover, in some cases of unilateral neglect the body schema continues to function in a normal manner. The patient can have normal motor control and can use the neglected side in various actions, despite the absence of a body image for that side (for discussion, see Gallagher 2005). This contrasts to cases of deafferentation in which the subject has lost aspects of the body schema and normal automatic control of movement, but retains an enhanced body image (Gallagher and Cole 1995).

Having or not having a body image is not the same as *not feeling* or *feeling* disembodied, respectively, since one may lack a body image and feel either weakly disembodied or completely disembodied (as in 3b), or not disembodied at all (as in unilateral neglect). Indeed, an explicit body image is not a constant feature of normal, everyday phenomenology, since in many cases, when we are involved in intentional action or some project that directs our attention to some object in the world, we lack any explicit experience of our own body—it becomes transparent, in just the way that Metzinger claims the self-model becomes transparent. In contrast to feeling disembodied, in these normal situations I have an implicit and pre-reflective sense of my embodiment, defined primarily in terms of motor or action capabilities (for convenience, I'll call this normal, everyday, implicit sense of embodiment "enactive embodiment").

Metzinger claims that "attentional and cognitive agency can *functionally* be decoupled from the process of autonomic self-regulation and the spatial self-representation necessary for generating motor behavior" (p. 499). This would seem to describe what Metzinger would call the functional disembodiment found in deafferentation. But Metzinger suggests that it is rather like the case of a virtual reality (VR) system which creates "the conscious experience of viewing one's own body as embedded in and interacting with a virtual world or the experience that there is a 'real you' not currently inhabiting your body" (p. 499). Clearly, however, in the VR case there would be no *functional* disconnection between one's real body and one's virtual body, since movement of the real body controls movement of the virtual one, although it may be that there is a phenomenal disconnect—that is, my body may seem to be somewhere where it is not, or I may have a very strong sense of being in the virtual body (see Cole, Sacks and Waterman 2000). Indeed, the same may be said about deafferentation, that is, that there is no functional disconnection, although for the opposite reason. In deafferented patients who have learned alternative methods of motor control, generated and sustained by attentional and cognitive agency, the body image (something like a virtual body) controls the movement of the real body. There is, in effect, an unusual (top-down), but nonetheless functional connection between the real body and attentional and cognitive agency. A truly functional (and specifically motoric and enactive) disembodiment does seem to exist in the cases of paralysis or possibly Anarchic Hand Syndrome, where a subject would have no motor control over his real body or body part.

Returning from these unusual cases to the more normal case of enactive embodiment, we note that motor control *per se* is not usually a matter of attentional or cognitive processes, if by that we mean attending to and cognitively controlling our bodily movements, or relying on a body image for motor control. Again, we normally have no explicit consciousness of motor control processes, most of which remain unconscious (see, e.g., Jeannerod 2003). Proprioceptive-kinaesthetic awareness is normally recessive and contributes to a pre-reflective sense of our enactive embodiment.

3. Being disembodied

Clearly, then, there are cases of phenomenal disembodiment without functional disembodiment (as in 3b). There are also cases of functional disembodiment without phenomenal disembodiment (for example, anosognosia in cases of paralysis, where the patient is paralyzed but seems to experience, or claims to experience, normal enactive embodiment). Are there any cases of *real* disembodiment — again, understanding the term 'real' as anything that has the ontological status bestowed by naturalism on things like brains, biosystems, and scientific equipment? Both phenomenal and functional disembodiment, of course, are real enough, but the brain in the vat scenario, (4), would take this a step further, involving the decoupling not only of attentional and cognitive agency, but also of all other aspects of experience, from the autonomic self-regulation of the real body, since, in this scenario, there is no real body. The wonderful thing about being a brain in a vat is that it would purportedly have no effect on our phenomenal embodiment. Phenomenally, the brain in the vat would not feel disembodied, assuming appropriate stimulation is maintained (and assuming the brain does not go Cotardian, or have, in this case, a veridical out-of-body experience). And functional motor control would even seem to work, that is, we would have a phenomenal sense of it working. Seemingly, it would make no difference in our experience whether we had a real body or not. Or would it?

What would allow us to experience the world exactly the way we do now, if we were brains in vats? What would have to happen to the vat brain in order to generate the experience that we normally have? Everything done to the vat brain would have to emulate what the body does. If visual experience were to be the same as embodied visual experience, then it would not do to simply attach a video feed. Apparently, the retina (which is part of the brain) would have to be stimulated in exactly the same way as an enactively embodied retina is stimulated. Proprioceptive-kinaesthetic input would have to be generated by direct sub-cortical stimulation, but with precisely the same calibrations provided by embodied proprioception. Likewise for activation of neuronal processes responsible for vestibular sense and balance. Motor commands would have to send signals to artificial transducers that would somehow activate mechanisms that cause the same effects in the world as my bodily movements, including the visual and kinaesthetic reafference associated with bodily movement, since, in this regard, experience normally depends on "a large number of different receptors external to the brain but internal to our body, which are, for instance, found in skin, muscles, joints, and viscera" (Metzinger p. 505).

We know that the body's very shape and design contribute directly to sensory "pre-processing," the specifics of which can be shown in studies of comparative anatomy. Auditory experience, for example, is constrained by the shape and location of the ears, which determine directional information by amplifying or filtering specific inputs (Chiel and Beer 1997). The brain's pre-motor and motor areas do not fully determine movement; rather, movement is re-engineered by the design of muscle and tendons, their degrees of flexibility, their geometric relationships to other muscles and joints, and their prior history of activation (Zajac 1993). Central planning of movement is always incomplete; it is part of a competitive system that requires what Andy Clark terms 'soft assembly'. The nervous system learns "to modulate parameters (such as stiffness [of limb or joint]) which

will then *interact* with intrinsic bodily and environmental constraints so as to yield desired outcomes" (Clark 1997, p. 45). Absent the body, such constraints would have to be artificially induced in the nervous system; and absent the complete nervous system, stimulation of brain areas would have to take the following principle into account: "The nervous system cannot process information that is not transduced by the periphery, nor can it command movements that are physically impossible for that periphery" (Chiel and Beer 1997: 554). This suggests that, short of a miraculous engineering project, at the most basic levels of perception and action the hypothetical experience of a brain in a vat would be essentially different from an embodied brain.

Of course these are the obvious things that pertain to our sensory-motor experience. Our experience is also colored by various emotions, moods, and feelings, and by other states not unconnected with processes that happen in various systems of the body. Just the right mix of hormones and neurotransmitters, together with properly measured, ongoing fluctuations in the level and placement of these chemicals would be required to generate the continually changing and richly colored experience that I have of the world. The difficult chemical engineering that the body accomplishes throughout our waking and sleeping hours means that the vat brain could not simply be immersed in a cocktail of chemicals. Metzinger, who describes the process of generating experience simply in terms of the activation of the minimally sufficient neural correlate, admits that this could generate only "a 'phenomenological snapshot'" of experience. And of course experience is not a snapshot, or even a series of snapshots. There is continuity and constant change, and these aspects of experience cannot be accounted for entirely in terms of the intentional content of what we experience. They also involve gut-level reactions and a sophisticated circulation of biochemicals.

Moreover, it is not simply a matter of stimulating neurons to create an explicit body image, since an explicit body image is not a necessary component of the experience of enactive embodiment. If there were no full-blown pre-reflective, biological, enactive lived body hooked up to this brain, then everything would have to be rigged in such a way that all of the embodied processes—the sensory-motor, the autonomic (sympathetic and parasympathetic), hormonal and neurotransmitter inputs—could be delivered at the right time and in the right way to the right part of the brain. The engineering project would be less like building Boston's "big dig" tunnel, and more like engineering the formation of the Grand Canyon on its original real-time geological timescale, and delivering its present day formation in all of its detail, including color, fauna and flora.

Let me be clear. Just as Metzinger's point is not that the brain in the vat scenario is possible, my point is not that the brain in the vat scenario is impossible (although it surely is). My point is that there is no such thing as *real* disembodied experience. Full-blown pre-reflective embodiment should not be thought of as simply a body image that could be generated by neuronal stimulation, or a simulated functional body-schematic system that could be activated in an exclusively neural matrix. The lived body, the body I live, is the real biological body, and if it were taken away, the life-support system that would have to replace it would necessarily be a real complex system that I could live and experience in the same way as I live and experience my body.

4. Metzinger's metaphors

As Metzinger says, "as agents, we live our body" (p. 296). And again, "it is the *body* which anchors us in reality—physically, and functionally, as well as phenomenally" (544). But if we think of ourselves as something different from our bodies, if we think of ourselves as simply a self-model, a "virtual self-model" (544), or an illusion generated by the brain, then our ontology is that of a ghost in a machine – a ghost generated by the machine. If *real* disembodiment is impossible, it is because there is nothing from which to subtract the body. This "nothing from which to subtract the body" can be read in two ways. *Either* there is no one there and what we call the self is an illusion, a model generated in the system; *or*, our bodies *are* us. To live my body means more than to be aware of my body or to have a body image. It means that I am my real body and in that facticity (because it is more than just a fact—it is a fact that I live through) I live in a real world. What I call my "self" is simply myself as a full-fledged embodied entity. What I experience as my self is a select subset of everything the body, which includes the brain, does. Moreover, my pre-reflective embodied self-consciousness is usually veridical. I am not a fiction or a ghost generated by the body. What I pre-reflectively experience of this existence is not an illusion, since my pre-reflective experience is itself generated by my real embodiment—by a brain that is part of a real bodily system, which is part of a real environmental system.

This doesn't mean that I experience myself exclusively *as* a body, or that things can't go wrong with the way that I conceive of and interpret myself. The natural engineering of the human body (including the reproductive and nurturing configurations, the shape of our hands, the structure of our faces, the brain and jaw structures that allow for language, and all of the functions that go along with all of these things, and which make us the rational, emotional, and social animals that we are) allows us to generate narratives and metaphors that lead us beyond the simple equation of self and body. I am this body, *and* I am more than this body. But in so far as this body, which includes my brain, allows me to be more than this body, through my actions and narratives and reasonings, the veracity and reality of who I am always comes back to this body. That there are real bodies means that there are real selves who experience themselves both as bodies and as more than bodies.

Metzinger disagrees with this. He wants to say that the self is not real ("no such things as selves exist in the world: Nobody ever *was* or *had* a self" [p. 1]). In effect, the self is a ghost generated by the machine. In making such claims, however, I think he may be getting tripped up and tangled up by his own metaphors. "In other, more metaphorical, words, the central claim of this book is that as you read these lines you constantly *confuse* yourself with the content of the self-model currently activated by your brain" (p. 1). It seems possible that as Metzinger wrote those lines his metaphor literally became confused with a theoretical model.

References

Chiel, H.J. & Beer, R.D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences*, 20, 553-557.

- Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.
- Cole, J., Sacks, O. & Waterman. I. (2000). On the immunity principle: A view from a robot. *Trends in Cognitive Science*, 4 (5), 167.
- Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Oxford University Press.
- Gallagher, S. & Cole, J. (1995). Body schema and body image in a deafferented subject. *Journal of Mind and Behavior*, 16, 369-390.
- Husserl, E. (1952). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*. Zweites Buch. M. Biemel (ed.), Husserliana IV. Haag: Martinus Nijhoff.
- Jeannerod, M. (2003). Self-generated actions. In S. Maasen, W. Prinz and G. Roth (Eds.), *Voluntary Action: Brains, Minds, and Sociality* (pp. 153-64). Oxford: Oxford University Press.
- Metzinger, T. (2004). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.
- Zajac, F. E. (1993). Muscle coordination of movement: A perspective. *Journal of Biomechanics*, 26, suppl 1, 109-124.