PSYCHE

What a Self Could Be

Marcello Ghin
Institute of Humanities, Philosophy
University of Paderborn
Warburgerstr. 100
33102 Paderborn
Germany

marcello.ghin@upb.de

Keywords: self, self-model theory of subjectivity, phenomenal self-model, being no one, self-sustaining system

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: Metzinger's claim that there are no such things as selves has given rise to a lot of discussions. By examining the notion of self used by Metzinger, I want to clarify what he means when saying that nobody ever was or had a self. Furthermore, I want to examine if there could be a notion of 'self' which is compatible with the Self-Model Theory of Subjectivity (SMT). I will argue that there is a notion of self which is not only compatible with the SMT, but that the SMT also provides the theoretical framework for developing such a notion.

1. Introduction

"Alice took up the fan and gloves, and, as the hall was very hot, she kept fanning herself all the time she went on talking: 'Dear, dear! How queer everything is to-day! And yesterday things went on just as usual. I wonder if I've been changed in the night? Let me think: was I the same when I got up this morning? I almost think I can remember feeling a little different. But if I'm not the same, the next question is, Who in the world am I? Ah, that's the great puzzle!'" (Lewis Carroll: Alice's Adventures in Wonderland, The Pool of Tears). Thomas Metzinger's book *Being No One* is a thrill. It is one of the best worked

out analyses of phenomenal self-consciousness, which have been published recently. It offers an answer to the "great puzzle": the Self-Model Theory of Subjectivity (SMT). The SMT describes Alice's experience of being someone in terms of the content of continuously updated dynamic representational processes, a phenomenal self-model (PSM). Alice's experiences are predicted by the SMT: the content of the PSM is highly flexible and can change from moment to moment. What is lacking is a stable core, a self in the traditional sense, understood as an ontological substance that could in principle exist all by itself, as a mysteriously unchanging essence that generates a sharp transtemporal identity for persons. Although I believe that there are good reasons for Metzinger's central *ontological* claim that there are no such things as selves, understood in the sense just mentioned, I believe that there are also good reasons for not entirely giving up the notion of self. After examining the notion of self used by Metzinger, I will try to argue that there is a notion of self which is based on and compatible with the SMT and that sticking to this notion of self is informative for scientific purposes in the sense that there are patterns in the world which can best be described as selves, and that human beings, among other entities, belong to this class of patterns.

2. Being no one? What a self is not

To approach the first goal, it might be helpful to have a second look at some quotes where Metzinger is explicit about his central ontological claim. Directly at the beginning, the reader is confronted with the following statement:

> […] Its main thesis is that no such things as selves exist in the world: Nobody ever *was* or *had* a self. All that ever existed were conscious self-models that could not be recognized *as* models. (Metzinger 2003: 1)

We find similar statements throughout the book:

> Please remember that one of the central metaphysical claims guiding this investigation is that no such things as selves exist in the world. (Metzinger 2003: 462)
>
> First, it is important to understand the central ontological claim: No such things as selves exist in the world. (Metzinger 2003: 563)
>
> No such things as selves or subjects of experience exist in the world. (Metzinger 2003: 577)
>
> No such things as selves exist in the world. (Metzinger 2003: 626)

This is striking, and this strong claim has without doubt produced much confusion. We experience ourselves as selves, and the book claims to be about consciousness, the phenomenal self-model, and the first-person perspective. And now we're told that we don't exist? Or is that really what we are told? It is probably worth mentioning that Metzinger does not say "no selves exist" but "no such things as selves" exist. That might just be a different way of saying the same thing, but it might also point towards a specific way in which we have to understand the central ontological claim. Clearly, Metzinger must have something in particular in mind when he says that no such *things* as selves exist in the world.

We get a first hint towards what Metzinger has in mind when saying no such *things* as selves exist already on the second page, when he speaks about the epistemic goal of his book:

> The epistemic goal of this book consists in finding out whether conscious experience, in particular the experience of *being someone*, resulting from the emergence of a phenomenal self, can be convincingly analyzed on subpersonal levels of description. A related second goal consists in finding out if, and how, our Cartesian intuitions – those deeply entrenched intuitions that tell us that the above-mentioned experience of being a subject and a rational individual can *never* be naturalized or reductively explained – are ultimately rooted in the deeper representational structure of our conscious minds. (Metzinger 2003:. 2)

The reference to "our Cartesian intuitions" opens space for interpreting the claim "that no such things as selves exist" as "no such things as selves, understood as in the sense of a Cartesian cogito, as a substance, exist". Indeed, this seems to be what Metzinger is after:

> This phenomenally transparent representation of invariance and continuity constitutes the intuitions that underlie many traditional philosophical fallacies concerning the existence of selves as process-independent individual entities, as ontological substances that could in principle exist all by themselves, and as mysteriously unchanging essences that generate a sharp transtemporal identity for persons. But at the end of this investigation we can clearly see how individuality (in terms of simplicity and indivisibility), substantiality (in terms of ontological autonomy), and essentiality (in terms of transtemporal sameness) are not properties of selves at all. At best, they are folk-phenomenological constructs, inadequately described conscious *simulations* of individuality, substantiality, and essentiality. And in *this* sense we truly are no one. (Metzinger 2003: 626)

So the central ontological claim is actually that no such things as selves exist, understood as "process-independent individual entities, as ontological substances that could in principle exist all by themselves, and as mysteriously unchanging essences that generate a sharp transtemporal identity for persons". Let us call that the *strong self*. As such, the central ontological claim is less challenging, and I guess that most researchers interested in consciousness and the self would agree. If this is what *being no one* means, i.e. that we are no one in the sense of this strong notion of self, I agree with Metzinger when he says that

> […] this first reading of the concept of "being no one" is only an answer to the rude traditional metaphysics of selfhood, and I think as such it is a rather trivial one. (Metzinger 2003: 626)

There is another reading of the claim that no such *things* as selves or subjects of experience exist, merely with the emphasis on 'things' in 'no such things'. When trying to develop an ontology, one can distinguish between continuants, things, and occurrents, events. Continuants, or things, are conceived of as having spatial parts, but no temporal parts. Occurents, or events, have spatial as well as temporal parts. Processes are a succession of events. A box of dynamite, in this sense, is a thing, an explosion a process. However, we can spell out things (continuants) in terms of processes, and we have reasons for believing that, ontologically speaking, all that exists are processes (Russel 1931, Quine 1960, Lewis 1986, Heller 1990). A 'thing', understood as something static, is not adequate for consciousness, the phenomenal self, and the first person perspective. This is clearly shown by Metzinger's multilevel analysis of the target properties of

consciousness, the phenomenal self, and the first-person perspective. Thinking about it in this way makes it clear why Metzinger substitutes "self" by "PSM":

> Metaphysically speaking, what we called "the self" in the past is neither an individual nor a substance, but the content of a transparent PSM. There is no unchanging essence, but a complex self-representational process that can be interestingly described on many different levels of analysis at the same time. For ontological purposes, "self" can therefore be substituted by "PSM". (Metzinger 2003: 626)

Being No One works with both readings, i.e. it shows why no such things as selves in the strong sense exist, and that selves could not be *things*. However, the strength of *Being No One* is not that it shows that our traditional concepts of 'self' don't work, but that the alternative that Metzinger offers explains why one might be tempted to think of oneself as a self in the strong sense.

3. Being someone? What a self could be

Now that we have seen what Metzinger has in mind when saying that no such things as selves exist, we can pursue the second goal, i.e. examine if there could be another notion of 'self' which is compatible with the self-model theory of subjectivity.

First of all, it is important to understand why our folk-psychological notion of self is misguided. It is based on a naïve realistic stance towards our conscious experiences. However Metzinger shows that the phenomenal self-model is not a self:

> The folk psychology of self-consciousness naively, successfully, and consequentially tells us that a self simply is whatever I subjectively *experience* as myself. (Metzinger 2003: 268)

This is due to the nature of the PSM, i.e. due to the fact that most parts of our PSM are transparent. On a surface level, it seems that we directly experience our own body, that we have introspective access to all our mental states and that our conscious experience forms a coherent, global whole. Due to the transparency of most of the content of our PSMs, we tend to believe that we are what we experience, not being aware of the fact that it is just a construct:

> 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. (Metzinger 2003:1)

However, believing that we are whatever we consciously experience as being ourselves would be committing "the error of phenomenological reification" (Metzinger 2003: p. 268). The main problem with phenomenological reification is that the content of the phenomenal self-model as such is not epistemically justified (Metzinger 2003: p. 404). Thus, Metzinger concludes that there is really no one having these experiences. To be

precise, there is no one who could confuse herself with the content of the phenomenal self-model:

> Do you recall how, in the first paragraph of the first chapter, I claimed that as you read these lines you constantly *confuse* yourself with the content of the self-model currently activated by your brain? We now know that this was only an introductory metaphor, because we can now see that this metaphor, if taken too literally, contains a logical mistake: There is no one *whose* illusion the conscious self could be, no one *who* is confusing herself with anything. (Metzinger 2003: 633-34)

According to Metzinger, that is why the answer to Alice's question "Who in the world am I", would be, sadly, "you are just an illusion, a hallucinatory content of an ongoing dynamic representational process." Alice does not have nor is a self. We all know that Alice existed first as part of the content of Lewis Carroll's phenomenal self-model and now continues to exist trough being integrated into our phenomenal self-models.

Lewis Carroll, however, just like any other living human being, might actually be in a better position than Alice. In the following I want to argue that human beings might actually not just be hallucinated selves, but real selves. Metzinger argues, using the metaphor of a neurophenomenological cave-man, that there is no one in the cave. However, the self-model theory of subjectivity does not constrain us to conceive of ourselves on the level of the contents of our conscious experience. Indeed, it presupposes, or at least makes the hypothesis very plausible, that we are biological organisms that construct the cave and the neurophenomenological caveman as a tool for orienting ourselves in the world. The notion of self that I want to suggest here and examine in the last part of this comment is, in a preliminary version, that a self is a metabolic self-sustaining system that operates under a functionally adequate phenomenal self-model (in which sense such a system is a self will be explained further down). The SMT provides the ground for such a notion of a self, as we can already find on the first page of *Being No One*:

> The phenomenal self is not a thing, but a process – and the subjective experience of *being someone* emerges if a conscious information-processing system operates under a transparent self-model. (Metzinger 2003: 1)

Saying that the subjective experience of *being someone* emerges if a conscious information-processing system operates under a transparent self-model is not sufficient for a philosophically interesting notion of a self. I think that we can enrich such a concept of a self by exchanging "conscious information-processing system" with "self-sustaining system". Without doubt, we can model self-sustaining systems as information-processing systems, but then we would miss an essential point.

The first step I want to take in order to show what we gain through the notion of self-sustainment and that the suggested notion is compatible with SMT is to look at Metzinger's positive ontological claims:

> All that, in an ontological sense, does exist are certain classes of information-processing systems operating under transparent self-models. For these systems, having such a self-model is just a new

> way of having access to themselves. Therefore *all* selves are either hallucinated (phenomenologically), or elements of an inaccurate, reificatory phenomenological descriptions. (Metzinger 2003: 462)

Three points are interesting here. We have already seen the first, stating that what exist are, according to the SMT, not selves in a strong sense, but information-processing systems that use transparent self-models. A new, and important point is that information-processing systems use transparent self-models to gain a new kind of access to themselves. As we will see, this is a crucial point for the possibility of a notion of self that is compatible with the SMT. The third point made in this statement is that it does not matter if the content of a PSM is a mere hallucination or a real representational content, there will always be a system realizing the phenomenal self-model. This point is again stressed in the section of *Being No One* on hallucinated selves:

> But as long as we hold on to a realist ontology, it will always remain true that *some* kind of physical system giving rise to the currently hallucinated self does exist. Again, it is important to note how the notion of a "hallucinated self" is not a contradiction in terms: what the hallucination is attributed to is not a conscious Cartesian ego, but simply the physical system as a whole. Just as it can generate a selfless phenomenal model of reality, the physical system as a whole can also hallucinate a self. (Metzinger 2003: 462)

Even though the SMT presupposes that there is always a system realizing the PSM, we have to be careful not to call just any system generating a PSM a self. As such, it is not clear at all why there should be information processing systems that generate conscious experience. The point that a transparent self-model is "just a new way of having access to themselves" can be seen as a first hint towards an answer to the question of why systems pay the high costs involved in generating conscious experience. Having access to themselves in a new way must have been advantageous in one or the other way. Looking at it from an evolutionary perspective, it seems plausible to say that those systems should be called selves that generate a transparent representational self-model, i.e. those systems that generate a phenomenal self-model that is adequate in the sense that operating under the phenomenal self-model enables the system to sustain itself. The SMT makes the hypothesis very plausible that we are biological organisms that operate under a phenomenal self-model with representational content. The content of our PSMs does not provide the basis for this claim, since it could always just be hallucinated content. However, the background assumptions on which the SMT rests and the adaptivity constraint make it plausible to assume that most contents of PSMs are representational content. The SMT *is*, first of all, a theory about human beings *in nonpathological* waking states, i.e. about biological organisms operating under a representational phenomenal self-model, not just any hallucinated phenomenal self-model:

> However, everywhere in this book where I am not explicitly concerned with this type of reality test, the following background assumption will always be made: the intended class of systems is formed by human beings in nonpathological waking states. (Metzinger 2003: 14)

This is what differentiates us as *selves* from *phenomenal selves* that emerged in organisms with identity disorders or from selves that emerge when actors enact other characters. We

don't say that persons with dissociative identity disorder or acted characters on stage are really a sum of different selves. In both cases, we are hesitant to call the phenomenal self a self, because it does not serve the function we think a self does in normal, non-pathological, non-artificial standard situations. Metzinger makes this explicit when speaking about the adaptivity constraint:

> If we want to understand how conscious experience, a phenomenal self, and a first-person perspective could be *acquired* in the course of millions of years of biological evolution, we must assume that our target phenomenon possesses a true teleofunctionalist description. (Metzinger 2003: 198)

As we can see here, it is important to understand that, when we want to know why conscious experience has emerged, why organisms pay this high metabolic price, having the consciousness experience of being someone serves the organism in specific ways. Metzinger does not provide us with a theory saying what the "true teleofunctionalist description" amounts to, but makes functionalism one of his background assumptions:

> I do not explicitly argue for teleofunctionalism in this book, but I will make it one of my implicit background assumptions from now on. (Metzinger 2003: 26)

But Metzinger is quite clear about what he thinks the function of consciousness is. Consciousness is the process by which information is made globally available (Baars 1988, 1997, Chalmers 1997) to the system for attention, action and cognition (not all kinds of global availability have to be given at each moment, and information can be made available for each kind in different degrees). Making information globally available for attention, action, and cognition serves as

> […] an instrument to generate successful behavior; like the nervous system itself it is a device that evolved for motor control and sensorimotor integration. Different forms of phenomenal content are answers to different problems which organisms where confronted with in the course of their evolution. Color vision solves another class of problems than the conscious experience of one's own emotion, because it makes another kind of information available for the flexible control of action. An especially useful way of illustrating this fact consists in describing phenomenal states as new *organs*, which are used to optimize sensorimotor integration of the information flow within a biosystem. (Metzinger 2003: 200-01)

Another way to put it is to conceive of self-modeling processes as instantiating "tools and weapons" (Metzinger 2003: p. 344) at many different levels in many different contexts. It helps us regulating a balanced homeostasis (or, to be more precise, a homeodynamics), avoid danger, enjoy pleasure or plan an international research project.

I already gave a preliminary notion of the concept of self that I want to outline above: a self is a self-sustaining system that operates under a phenomenal self-model. There are a couple of advantages, which we gain from constraining our notion to self-sustaining systems and not just any system that generates a phenomenal self-model. This notion enables us to understand why something like selves emerged. Furthermore, it justifies some of our intuitions about what we are. For example, saying that human beings

in non-pathological waking states are selves in the sense of a self-sustaining system operating under a PSM justifies our experience of ourselves as constituting a unified self over time because biological organisms maintain *gene identity* (not to be confused with gene identity from descent) (Lewin 1922, 1923, Armstrong 1980). All we have to do is to be cautious. A unified self is not a transtemporal identity in the sense of an unchangeable essence, it is a process. The process we are looking for is that of a metabolic self-sustaining process (Jordan & Ghin, forthcoming). It seems safe to say that the kind of conscious systems, which experience themselves as selves, that we know, are biological systems, i.e. embodied embedded biological organisms, that are metabolic self-sustaining systems. The notion of self-sustainment enables us to flesh out what it means for a system to be embodied. For a first approximation, we can say that the notion of self-sustainment basically serves the same purpose as Varela's notion of autopoiesis. The Greek concept autopoiesis means self-producing, and Varela uses it for giving a definition of living systems:

> An autopoietic system is organized (defined as a unity) as a network of processes of production (synthesis and destruction) of components such that these components:
> (i) continuously regenerate and realize the network that produces them, and
> (ii) constitute the system as a distinguishable unity in the domain in which they exist. (Varela 1992: 5)

The problem is that, according to this abstract notion of autopoiesis, a civilization is an autopoietic system, and thus is not very helpful for the current purpose. However, if we enrich the notion of autopoiesis by adding metabolism in the sense that the system creates and maintains its own body, we gain a notion of embodiment that allows us to say that, already on this level, the system forms a distinct unit in the sense that its self-sustaining process creates a self/world distinction for the system itself. However, primitive self-sustaining systems generate a self-world distinction for themselves without being aware of themselves as selves, and would thus not yet provide a psychologically interesting notion of self. Once self-sustaining organisms create, under evolutionary pressure, a specific level of complexity, they will have to create a self-model in order to control themselves (i.e. produce coherent actions), and will produce those representational structures that give rise to conscious experiences, the most primitive form of which will likely be that of an emotional self reflecting and evaluating changes of organism-environment interactions, experienced as a feeling (Damasio 1999, Prinz 2004), until we finally arrive at the most complex form we know of, fully fledged human cognitive agents that even develop theories about concepts like "self" to refer to themselves. If we understand ourselves as such systems, it becomes clear that referring to ourselves as selves is not just an illusion, but that there really are patterns in the world (the dynamic self-sustaining patterns of biological organisms) onto which we can map the concept.

One could raise the question why we should stick to the notion of self and not just speak about "self-sustaining systems + phenomenal self-models." It is important to keep in mind that a reductionist theory of the self does not imply that we have to eliminate *selves* from our ontology. As Bermúdez (1997) argues, reducing one theory to another can also provide legitimacy to the reduced theory in the sense that it shows how the reduced theory fits within the realm of the other, and can thus be seen as being validated

through the reduction. The self-model theory of subjectivity clearly shows how such a reduction can be done for the notion of a *self*. But, rather than saying that a successful reductionist theory of phenomenal self-consciousness and the first-person perspective shows that we could, in principle, drop the notion of a *self*, we can see it as enabling us to understand how *selves* fit into the world.

Self is a concept which is not only used in folk-psychology but also in many different scientific theories. Admittedly there have been many intuitions about what the notion of self amounts to, and many arguments about the status of a self as forming an irreducible kind of entity, all of which, from the perspective of the self-model theory of subjectivity, can no longer be defended. But rather than eliminating the notion and questioning the explanatory power of theories that appeal to selves, we could say that we now found a way to legitimate the claims made by these theories by providing a framework which unifies higher- and lower level theories about phenomenal self-consciousness and the first-person perspective.

In this sense, I believe that there is a notion of self that is not only compatible with the SMT, but also proves to be helpful for scientific purposes as it picks out a distinct class of information-processing systems with unique features, the main feature would be that these systems are able to experience themselves as selves and thus understand themselves as the authors of their actions. We, as human beings, are part of this class.

Self-sustaining systems come in degrees, from single-cellular organisms, where self-sustainment is restricted to the production and maintenance of the cellular structure, up to human beings, where the system tries not only to maintain its body, but also a coherent model of itself with specific abstract characteristics, like being altruistic, creative, funny or attractive. I don't want to offer a full-fleshed theory of the concept self based on the notions of self-sustainment and the self-model theory of subjectivity here. It will remain open for discussion, where we want to draw a line and apply the notion of self. Human beings are good candidates, other primates seem to be as well, but an amoeba might not be. Furthermore, it is open if we apply the notion of self only to metabolic self-sustaining systems, i.e. if it is possible that there could be non-biological systems that operate under functional adequate phenomenal self-models in the sense that the PSMs enable them to maintain a coherent model of themselves. This possibility cannot be ruled out at the moment, and thus saying that a self is an autocatalytic self-sustaining system operating under a PSM is only speaking about sufficient constraints a system has to satisfy in order to count as a self. One advantage is that the notion of an autocatalytic self-sustaining system grounds content and valence (Jordan & Ghin, forthcoming). So far we cannot see how content or valence can be grounded for non-biological systems. The role of the notion of self as developed here can thus best be seen as providing a basis for other theories.

The two goals of the comment where to clarify what exactly Metzinger had in mind when saying nobody ever was or had a self, and to examine if there could be another notion of 'self' which is compatible with the Self-Model theory of subjectivity. To summarize, *Being No One* is not only compatible with a notion of a self understood as a process, but also provides a theoretical framework for developing such a notion. The self-model theory of subjectivity makes it possible to argue that we are no one at one level and that we are someone at another level.

References

Armstrong, D. (1980). Identity Through Time. In: P. van Inwagen (ed.), *Time and Cause*. Dodrecht, 1980.

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.

Bermúdez, J. L. (1997). Reduction and the self. *Journal of Consciousness Studies*, **4**, pp. 458-66.

Carroll, L. C. (1866). *Alice's Adventures in Wonderland. The Pool of Tears*. Reprinted in Carroll, L., C.: *The Complete, Fully Illustrated Works* (1995[2]), New York: Gramercy Books

Chalmers, D. J. (1997). Availability: The cognitive basis of experience? *Behavioral and Brain Sciences*, **20**, pp. 148-9.

Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.

Heller, M. (1990). *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*. Cambridge: Cambridge University Press.

Jordan, J. S., & Ghin, M. (forthcoming). Born to be Wild.

Lewin, K. (1922). *Der Begrifff der Genese in Physik, Biologie und Entwicklungsgeschichte. Eine Untersuchung zur vergleichenden Wissenschaftslehre*. Berlin.

Lewin, K. (1923). Die zeitliche Geneseordnung. *Zeitschrift für Physik*, **13**, pp. 62-81.

Lewis, D. (1986): *On the Plurality of Worlds*. Oxford: Blackwell.

arcourt Brace and Company.

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

Prinz. J. J. (2004). *Gut Reactions. A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

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

Russel, B. (1931): *The ABC of Relativity*. London: Keagan Paul.

Varela, F. J. (1992). Autopoiesis and a Biology of Intentionality. In: McMullin & Murphy (eds.) *Autopoiesis and Perception: A Workshop with ESPRIT BRA 3353* (pp. 4-14).