Self models

Post-publication activity

Curator: Thomas Metzinger

Dr. Thomas Metzinger, University of Mainz, Germany

The self-model theory of subjectivity (SMT)

The concept of a **self-model** plays the central role in a philosophical theory of consciousness, the phenomenal self and the first-person perspective. This specific theory is the so-called "self-model theory of subjectivity" (**SMT**; see Metzinger 2003a, 2005a). However, SMT is not only a conceptual framework in analytical philosophy of mind, but at the same time an interdisciplinary research program spanning many disciplines from neuroscience, cognitive science, neuropsychology and psychiatry to artificial intelligence and evolutionary robotics (e.g., Blanke & Metzinger 2009, Metzinger 2007, Lenggenhager et al. 2007, Windt & Metzinger 2007, Metzinger & Gallese 2003). The theory simultaneously operates on phenomenological, representational, functional and neuroscientific levels of description, using a method of interdisciplinary constraint satisfaction (see Weisberg 2007, section 2, for critical discussion). The central questions motivating the SMT are: How, in principle, could

a consciously experienced self and a genuine first-person perspective emerge in a given information-processing system? At what point in the actual natural evolution of nervous systems on our planet did explicit self-models first appear? What exactly made the transition from unconscious to conscious self-models possible? Which types of self-models can be implemented or evolved in artificial systems? What are the ethical implications of machine models of subjectivity and self-consciousness? What is the minimally sufficient neural correlate of phenomenal self-consciousness in the human brain? Which layers of the human self-model possess necessary social correlates for their development, and which ones don't? The fundamental question on the conceptual level is: What are the necessary and sufficient conditions for the appearance of a phenomenal self?

The phenomenal self-model (PSM)

The core notion of the theory is the concept of a "phenomenal self-model" (**PSM**). The PSM is that partition of an organism's self-model which is conscious because it satisfies an additional set of functional constraints like, for example, availability for introspective attention and for selective, flexible motor control. What exactly the necessary and sufficient conditions for phenomenality actually are, is, of course, presently still unknown. SMT (Metzinger 2003a, chapter 3) offers a set of ten such potential constraints, the two most important

of which are functional integration with a system's "virtual window of presence" (i.e., its internal representation of time and an extended "Now") plus ongoing, dynamical integration into a single, overarching world model (i.e., a multimodal representation of the current situation as a whole). A "model" is a working concept for a type of mental representation that does not contain variables, represents a situation by structural correspondence (i.e., via a partial relational homomorphy), the elements of which can correspond to perceptible entities as well as to abstract notions, which can be taken offline for internal "dry runs", and which may or may not satisfy the constraints for becoming a *phenomenal* mental model (see Craik 1943, 51pp; Johnson-Laird, 1983, 1989, p. 488; Metzinger 2003a, section 3.3 for more on the history of the concept). What we call “the conscious self” in folk-psychological discourse is a specific form of representational content, characterized by specific functional properties, which in turn can be physically realized in a large number of different ways. The *phenomenal* content of the PSM (how its content is subjectively experienced) locally supervenes, i.e., it is fixed as soon as all contemporaneous and internal properties are fixed. For example, as soon as all of the relevant parameters in the brain are set, the phenomenology going along with the activation of a PSM is determined as well. This may not be true of its *intentional* content, because self-knowledge (as opposed to subjective, self-*experience*) is arguably co-determined

by external factors like the way in which the system is historically and socially situated. One background assumption is that for every biological PSM, in the species known to us, there exists a minimally sufficient neural correlate of self-consciousness (NCSC; see Metzinger 2000). This NCSC would be a specific set of neurofunctional properties of which it is true that (1) it reliably activates a phenomenal self and that (2) it possesses no proper subset that suffices for the corresponding states of consciousness. The NCSC defines the set of properties that are *relevant* for a scientific explanation of phenomenal self-consciousness. If the predicates describing the relevant, locally supervening phenomenal properties and those referring to the neurofunctional properties determining them are nomologically coextensive, a reductive identification of the PSM is possible. The PSM would then be simply identical to the NCSC.

The self-model theory aims at maximally parsimonious framework for the scientific investigation of self-consciousness: There is no such thing as a substantial self (as a distinct ontological entity, which could in principle exist by itself), but only a dynamic, ongoing *process* creating very specific representational and functional properties. Self-consciousness is a form of physically realized representational content. This metatheoretical framework has implications for many of the first-order empirical sciences of the mind. For

example, the methodological core of scientific psychology can now be analyzed in a clearer, fresher and more fruitful way. Psychology, from this perspective, is self-model research: It is the scientific discipline that focuses on the representational content, the functional profile, and the neurobiological realization of the human self-model, including its evolutionary history and its necessary social conditions. Psychiatric disorders can be described and diagnosed more systematically, e.g. as specific representational contents in the self-model that have been lost, hyperactivated, or decontextualized, which are not available for introspective access any more, or which have become causally autonomous and functionally dissociated in other ways (Metzinger 2003a: ch. 7; 2004c)

Aside from the representational level of description, one can also develop a *functional analysis* of the self-model. Whereas representational states are individuated by their content, a functional state is conceptually characterized by its *causal role*: the causal relationships it bears to input states, output states, and other internal states. An active self-model then can be seen as a subpersonal functional state: a discrete set of causal relations of varying complexity that may or may not be realized at a given point in time. Since this functional state is realized by a concrete neurobiological state, it plays a certain causal role for the system. For instance, it can be an element in an information-processing account. The perspective of

classic cognitive science can help illustrate this point: the self-model is a transient computational module that is episodically activated by the system in order to control its interactions with the environment. In other words, what happens when you wake up in the morning, i.e., when the system that you are "comes to itself," is that this transient computational module, the PSM, is activated – the moment of "waking up" is exactly the moment in which this new instrument of intelligent information-processing emerges in your brain. It does so because you now need a conscious self-model in order to achieve sensorimotor integration, generate complex, flexible and adaptive behavior, and attend to and control your body as a whole. The assumption in the background is that the space of consciousness essentially is the *space of availability for selective, high-level attention*, and that the activation of a *conscious* self model becomes necessary whenever an organism (or an artificial system), in order to solve a certain task, needs an integrated internal representation of certain of its own *global* properties to make them available for selective resource allocation and deeper processing. This idea, namely, that the PSM is that partition of the currently active self model functionally characterized by availability (but not necessarily ongoing access), is a central working hypothesis under SMT, but its truth has not yet been empirically demonstrated.

For higher forms of intelligence, in order to integrate higher levels of behavioral and cognitive complexity, a

system needs a coherent self-representation, a consistent internal model of itself as a whole. A PSM is an instrument for global control. In our own case, the conscious self-model is an episodically active representational entity whose content is determined by the system's very own properties. Interestingly, we do not have a PSM in dreamless deep sleep, but there certainly is a PSM in the dream state. It can be characterized as an instrument for interacting with an exclusively internal environment, exhibiting not only a unique phenomenological profile, but a corresponding set of typical representational, functional and neurobiological features which can be described in a fine-grained manner. (Windt & Metzinger 2007). Many altered and pathological states of consciousness can be analyzed as deviant forms of self-modeling (Metzinger 2003, chapter 7). The evolution of higher cognition generally can be described as an increasing ability to take the PSM offline and to develop more abstract forms of mental-self simulation. For instance, simulating past or possible future states of the situated organism as a whole while comparing them to the current online self-model will enable memory, learning, or planning. Abstract thought would then have evolved out of the fundamental capacity to create forward models and to internally simulate elementary motor behavior and its perceptual consequences (Cruse 2007).

The development of ever more efficient self-models as a new form of „virtual organ" is also a precondition for the

emergence of complex societies. Plastic and ever more complex self-models not only allowed somatosensory, perceptual, and cognitive functions to be continuously optimized, but also made the development of social cognition and cooperative behavior possible. The most prominent example, of course, is the human mirror system, a part of our unconscious self-model that resonates with the self-models of other agents in the environment through a complex process of motor-emulation - of "embodied simulation," as Vittorio Gallese (2005) aptly puts it – for example, whenever we observe goal-directed behavior in our environment. Such mutually coupled self-models, in turn, are the fundamental representational resource for taking another person's perspective, for empathy and the sense of responsibility, but also for metacognitive achievements like the development of a concept of self and a theory of mind (for possible neurobiological correlates of these basic social skills, see Gallese & Goldman 1998, Metzinger & Gallese 2003).

The obvious fact that the development of our self-model has a long biological, evolutionary, and (a somewhat shorter) social history can now be accounted for by introducing a suitable version of teleofunctionalism as a background assumption. The development and activation of this computational module plays a role *for* the system: the functional self-model possesses a true evolutionary description, metaphorically speaking it was a weapon that

was invented and continuously optimized in the course of a “cognitive arms race” (Clark 1989: 61) The functional basis for instantiating the phenomenal first-person perspective can be seen as a specific cognitive achievement: the ability to use a centered representational space. In other words, phenomenal subjectivity (the development of a subsymbolic, non-conceptual first-person perspective) is a property that is only instantiated when the respective system activates a coherent self-model and integrates it into its global world-model.

Self-models in machines

Self-models can be entirely unconscious, and they can be realized in artificial systems as well (an excellent recent study is Schilling & Cruse, 2012). In 2006, Josh Bongard, Victor Zykov and Hod Lipson have created an artificial “starfish“ that gradually develops an explicit internal self-model. Their four-legged machine uses actuation-sensation relationships to indirectly infer its own structure and then uses this self-model to generate forward locomotion. When part of its leg is removed, it adapts its self-model and generates alternative gaits – it learns to limp. In other words: It is able to restructure its body-representation following the loss of a limb. It can learn. This concept may not only help develop more robust machines and shed light on the phylogeny and the ontogeny of self-modeling in animals, but is also

theoretically interesting, because it demonstrates for the first time that a physical system has the ability, as the authors put it, to “autonomously recover its own topology with little prior knowledge” by constantly optimizing the parameters of its own resulting self-model.

As we see, the robot initially performs an arbitrary motor action and records the resulting sensory data. The model synthesis component then synthesizes a set of 15 candidate self-models using stochastic optimization to explain the observed sensory-actuation relationship. The robot then synthesizes an exploratory motor action that causes maximum disagreement among the different predictions of these competing self-models. This action is physically carried out, and the 15 candidate self-models are subsequently improved using the new data. When the models converge, the most accurate model is used by the behavior synthesis component to create a desired behavior that can then be executed by the robot. If the robot detects unexpected sensor-motor patterns or an external signal resulting from unanticipated morphological change, it reinitiates the alternating cycle of modeling and exploratory actions to produce new models reflecting this change. The most accurate of these new models is then used to generate compensatory behavior and recover functionality.

What can be learned from this example? First, a self-model can be entirely unconscious; that is, it can be seen as the product of an automatic “bottom-up” process of

dynamical self-organization; second, it is not a "thing" (or a model of a thing) at all, but based on a continuous, ongoing modeling process; third, it can exhibit considerable plasticity (i.e., it can be modified through learning); and fourth, in its origins it is not based on language or conceptual thought, but very likely on an attempt to organize global motor behavior. More precisely, a body-model has the function of integrating sensory impressions with motor output in a more intelligent and flexible manner. The unconscious precursor of the PSM clearly was a new form of intelligence.

You do not have to be a living being in order to have a self-model. Non-biological SMT-systems are possible. As a self-model can be entirely unconscious, it does not have to be a phenomenal self-model (i.e., a PSM). Awareness obviously is a second step (see Metzinger 2003a, section 3.2, for an additional set of ten constraints to be satisfied for conscious experience). A self-model supports fast learning processes in a number of different ways, clearly making a system more resilient and intelligent. Fourth, it is a virtual model or "virtual organ" above, and one of its major functions consists in *appropriating* a body by using a global morphological model to control it as a whole. The technical term for this type of global self-control is "second-order embodiment" (Metzinger 2006b; see also Schilling & Cruse 2008, 2012). One of the core intuitions behind SMT is that a self-model allows a physical system to "enslave" its low-level dynamics with the help of a

single, integrated, and internal whole-system model, thereby controlling and functionally "owning" it. This is the decisive first step towards becoming an autonomous agent.

First-order embodiment, second-order embodiment, and third-order embodiment

The term "embodiment" has become almost semantically vacuous, because it is now used in many different academic disciplines and very different research contexts. The concept of a self-model permits the introduction of an interesting conceptual clarification.

"First-order embodiment" (**1E**) is aimed at and can be found, for instance, in biorobotics and in all "bottom-up approaches" to artificial intelligence. The basic idea is to investigate how intelligent behavior and other complex system properties, which we previously termed "mental," can naturally evolve out of the dynamical, self-organizing interactions between the environment and a purely physical, reactive system that does not possess anything like a central processor or "software" and exhibits no explicit computation. For researchers in 1E, the relevant questions are: How could the very first forms of pre-rational intelligence emerge in a physical universe? How could we acquire a flexible, evolvable, and coherent behavioral profile in the course of natural evolution? How is it possible to generate intelligent behavior without

explicit computation? Here is an example of 1E, the tripod gait as exhibited by the walking machine Tarry II:

"Second-order embodiment" (**2E**) can develop in a system that satisfies the following three conditions: (a) we can successfully understand the intelligence of its behavior and other "mental" properties by describing it as a *representational* system, (b) this system has a single, explicit and coherent self-representation of itself *as being an embodied agent*, and (c) the way in which this system uses this explicit internal model of itself as an entity possessing and controlling a body helps us understand its intelligence and its psychology in *functional* terms. Some advanced robots, many primitive animals on our planet, and possibly sleepwalking human beings or patients during certain epileptic absence seizures (as discussed in Metzinger 2003a) could be examples of 2E. Above, when discussing the Starfish, we have just encountered a model for the second working concept. The starfish-robot walks by using an explicit, internal self-model, which it has autonomously developed and which it continuously optimizes. It dynamically represents itself as an embodied, on an entirely non-conceptual, non-propositional, and unconscious level.

As we see, the robot initially performs an arbitrary motor action and records the resulting sensory data. The model synthesis component then synthesizes a set of 15 candidate self-models using stochastic optimization, to explain the observed sensory-actuation causal

relationship. The robot then synthesizes an exploratory motor action that causes the most disagreement among predictions of these competing self-models. That action is physically carried out, and the 15 candidate self-models are improved using the new data. When the models converge, the most accurate model is used by the behavior synthesis component to create a desired behavior that can then be executed by the robot. If the robot detects unexpected sensor-motor patterns or an external signal as a result of unanticipated morphological change, the robot reinitiates the alternating cycle of modelling and exploratory actions to produce new models reflecting the change. The new most accurate model is now used to generate a new, compensating behavior to recover functionality.

What is "third-order embodiment" (‘’’3E’’’)? 3E is the rare and special case in which a physical system not only explicitly models itself as an embodied being, but also maps some of the representational content generated in this process directly onto conscious experience. 3E-systems possess true phenomenological descriptions. That is, 3E means that in addition, you consciously *experience* yourself as embodied, that you possess phenomenal self-model (PSM). Human beings in ordinary wake states, but also orangutans swinging from branch to branch at great height, could be examples of 3E: they have an online model of their own body as a whole that has been elevated to the level of global availability and

integrated within a virtual window of presence. They are consciously present as bodily selves.

It is important to clearly differentiate between bodily self-knowledge and bodily self-consciousness. The *phenomenal* content of the bodily self model locally supervenes, it is fully determined as soon as all internal and contemporaneous properties of the central nervous system are fixed. For example, human beings are able to experience themselves as fully embodied selves, even when input from the physical boy is minimal, for instance during nocturnal dreams, when the brain is in an offline situation. A brain in a vat, possessing the right kind of microfunctional architecture, could therefore realize the *phenomenology* of 3E without possessing any kind of self-knowledge or *intentional* content in the true sense of the word. Brains in vats, however, do not belong to the intended class of systems, they are not the explanatory target of SMT. The question if 3E could exist in the absence of 1E and 2E is of a purely conceptual nature, as the interesting class of systems is constituted by those systems possessing 3E *because* they already have 2E an 1E as enabling conditions, plus the right kind of evolutionary history.

Here is an empirical example, investigating the phenomenology of phantom limbs. How “ghostly” are phantom limbs? Can one measure the “realness” of the PSM? A recent case-study by Brugger and colleagues introduced a vividness rating on a 7-point scale, which

showed highly consistent judgments across sessions for their subject AZ, a 44-year-old university educated woman born without forearms and legs. As long as she remembers, she has experienced mental images of forearms (including fingers) and legs (with feet and first and fifth toes) – but, as the figure below shows, not as realistic as the content of her non-hallucinatory PSM. Functional magnetic resonance imaging of phantom hand movements showed no activation of primary sensorimotor areas, but of premotor and parietal cortex bilaterally. Transcranial magnetic stimulation of the sensorimotor cortex consistently elicited phantom sensations in the contralateral fingers and hand. In addition, premotor and parietal stimulation evoked similar phantom sensations, albeit in the absence of motor evoked potentials in the stump. These data clearly demonstrate how body parts that have never been physically developed can be phenomenally simulated in sensory and motor cortical areas. Are they components of an innate body model? Or could they have been "mirrored into" the patient's self-model through the visual observation of other human beings moving around?

The general framework emerging from this threefold distinction is that human beings permanently possess 1E and 2E: a considerable part of our own behavioral intelligence is achieved without explicit computation and results directly from physical properties of our bodies, such as the genetically determined elasticity of muscles

and tendons, or the degrees of freedom realized by the special shape of our joints. Moreover, certain parts of our unconscious self-model, such as the immune system and the elementary bioregulatory processes in the upper brain stem and the hypothalamus, are continuously active. Two other input layers creating a permanent functional link into the physical organism itself are the vestibular system and continuous interoceptive self-preception (Seth, Suzuki & Critchley, 2012). Another candidate for an important aspect of the unconscious self-model, a representation of global properties of the body, is the body schema. Having an unconscious body schema is clearly a new, biological form of intelligence: having a body schema means having 2E. Only episodically, during wakefulness and in the dream state, do human beings realize 3E.

The relationship between 2E and 3E can be investigated by experimentally manipulating the phenomenal property of "body ownership" through multisensory conflict (e.g., Lenggenhager et al. 2007; see Blanke 2012 for review). Global body ownership is the conscious experience that a perceived body is *your* body, that the self is spatially located within its borders. Body ownership is theoretically relevant, because it is a good candidate for the most *simple* form of self-awareness - without agency, volition, or explicit conscious cognition (Blanke & Metzinger 2009). Recent data demonstrate that during multisensory conflict, participants felt as if a virtual body seen in front of them was their own body and mislocalized themselves

toward the virtual body, to a position outside their bodily borders. Experimentally creating illusions of the globalized, multisensory awareness of selfhood in a controlled manner with virtual reality technology opens a new avenue for the investigation of the neurobiological, functional, and representational aspects of embodied self-consciousness.

The transparency constraint

Many, but not all, conscious representations are "phenomenally transparent". **Transparency** in this sense is a property of phenomenal representations only: Unconscious representations are neither transparent nor opaque. It is a phenomenological concept, and not an epistemological one.

A broad standard definition of phenomenal transparency is that it essentially consists in only the content properties of a conscious mental representation being available for introspection, but not its non-intentional or "vehicle-properties". Typically, it will be assumed that transparency in this sense is a property of all phenomenal states. However, this definition is unsatisfactory, because it violates important phenomenological constraints: Introspective unavailability of the construction process is not a necessary condition for phenomenality, as non-intentional and vehicle properties frequently *are* accessible for introspection. Not all phenomenal states are transparent. Transparency comes in degrees.

Transparency holds if earlier processing stages are unavailable for attentional processing. Transparency results from a structural/functional property of the neural information processing going on in our brains, which makes earlier processing stages attentionally unavailable.

An example for transparency in the external world model is successful sensory perception: The apple in my hand is perceived not as a form of representational content, but simply as an object in the environment. It is as if I would "see through" my own internal representation and as if I was directly and immediately in contact with the environment. The phenomenology of transparency is the phenomenology of naïve realism. An example for opacity in the external world model would be pseudo-hallucinations, like the breathing, abstract geometrical patterns in a visual pseudo-hallucination. The dynamically evolving, consciously experienced patterns are earlier processing stages in the visual cortex, perhaps a process of disambiguating between a number of possible interpretations and settling into a stable state. This immediately takes the naïve realism away, because the system introspectively discovers that what is going on is actually a *representational* process – which takes place in a medium and might be a misrepresentation of the actual state of affairs. In the PSM, the standard example for transparency is the body image: It is experienced not as an internal construct, but as something irrevocably real. The phenomenology of conscious thought, however, is

characterized by opacity: While engaging in cognitive operations, we always experience ourselves as operating with mental *representations*, something that is internal, self-constructed, and might well be false, a misrepresentation.

The central claim of SMT is that a **phenomenal self** emerges if and only if a system operates under a transparent PSM. The phenomenal property of selfhood is instantiated whenever a system has a conscious self-model that it cannot introspectively recognize *as* an internal model. The phenomenological prediction is that if a system's PSM became entirely opaque, the conscious experience of selfhood would disappear. SMT also points out that what makes the brains of human beings special in comparison to all other animals on the planet is that their PSM possesses a very large opaque partition. Human beings have a conscious self-model that allows them to subjectively experience themselves *as representational systems* to a degree which was so far unprecedented.

The phenomenal model of the intentionality relation (PMIR)

Phenomenologically, a transparent world-model gives rise to a reality. A transparent system-model gives rise to a phenomenal self that is embedded in this reality. If there is also a transparent model of the transient and constantly changing relations between the perceiving, acting self and objects, goal-states and persons in this reality, this results

in a phenomenally experienced **first-person perspective**. SMT draws attention to a point, which has been frequently overlooked in the past: The classical intentionality-relation (Brentano 1874; see Metzinger 2006 for an application to conscious volition) can *itself* form the content of a conscious mental representation. In beings like us, there exists a phenomenal model *of* the intentionality relation (**PMIR**). We have, as it were, the capacity to "catch ourselves in the act": At times we have higher-order conscious representations of ourselves *as* representing. On the other hand, from an empirical point of view, it is highly plausible to assume that many nonhuman animals are intentional systems, but that their nervous systems do not allow them to ever become aware of this fact.

The central idea is that the intentionality-relation can itself be a form of phenomenal content, and that this will typically be a form of dynamical, subsymbolic content (as for example in the subjective experience of attending to a visual object or of being currently directed at an action goal). Only rarely will this take place on the level of explicit, high-level cognition. The core idea behind the theoretical notion of a **PMIR** is that human beings do not only represent individual objects or goal states, but that in many representational acts we also co-represent the *representational relation* itself – and that this fact is relevant for understanding what it means that the contents of consciousness is experienced as involving a "first-person perspective."

A genuine inner perspective arises if only and only if the system represents itself as currently interacting with the world to itself, and if it does not recognize this representation *as* a representation. A phenomenal first-person perspective is a transparent conscious model of the intentionality relation, a **transparent PMIR**. It represents itself as directed towards certain aspects of the world. Its phenomenal space is a *perspectival* space, and, phenomenologically, its experiences now are *subjective* experiences. The existence of a stable self-model allows for the development of the "perspectivalness of consciousness": the existence of a single, coherent, and temporally stable reality-model that is representationally centered in a single, coherent, and temporally stable phenomenal subject, a model of the system *in the act of experiencing*. A conscious human being is an example of a system that is capable of dynamically co-representing the representational relation while representational acts are taking place, and the instrument it uses for this purpose is the PMIR. The PMIR, the phenomenal model of the intentionality relation, is just another naturally evolved virtual organ, just like the PSM. The content of higher-order forms of self-consciousness is always relational: the self *in the act of knowing* (Damasio 1999: 168ff), the *currently acting* self. The ability to co-represent this intentional relationship itself while actively constructing it in interacting with a world is what it means to be a subject. In standard situations, the consciously experienced first-person perspective is the

content of a transparent PMIR.

The possession of a PMIR brings about new functional properties. A PMIR allows for a dynamical representation of transient subject-object relations, and thereby makes a new class of facts globally available. For instance, a PMIR allows a representational system to discover the fact that it *is* a representational system, that some of its internal states possess content. It also allows a system to discover the fact that it *has* attentional mechanisms, that it is capable of flexible internal resource allocation. A PMIR allows an agent to consciously experience the fact that it is *directed at goal-states*, or at the mental state of another agent in the environment, and so on.

A PMIR makes these specific types of information globally available within a virtual window of presence. Globally available information enables selective and flexible control of behavior. The general picture emerging is that phenomenal mental models, states of consciousness, are instruments used to make a certain subset of information currently active in the system globally available for the control of action, for focal attention and for cognitive processing. To have a conscious first-person perspective, to activate a transparent PMIR, makes a large amount of *new* information available for the system, and this could have been one of the reasons why it proved to be adaptive. If the activation of a PMIR takes place within an integrated, transparent world-model functionally integrated with a transparent virtual window of presence,

then this will result in the phenomenology of a knowing subject, which is currently present in a world.

References

Blanke, O. & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. Trends in Congitive Sciences (13:1), 7-13. doi:10.1016/j.tics.2008.10.003.

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. Nature Reviews Neuroscience (13:8), 556–571. doi:10.1038/nrn3292.

Brentano, F. (1973)[1874]. Psychologie vom empirischen Standpunkt. Erster Band. Hamburg: Meiner.

Clark, A. (1989). Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing. Cambridge, MA: MIT-Press.

Craik, K.J.W. (1943). The Nature of Explanation. Cambridge: Cambridge University Press.

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

Gallese, V. (2005). Embodied simulation: from neurons to phenomenal experience. Phenomenology and the Cognitive Sciences 4: pp. 23-38. doi:10.1007/s11097-005-4737-z.

Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. Trends in Cognitive Sciences 2: 493-501. doi:10.1016/S1364-6613(98)01262-5.

Johnson-Laird, P.N. (1983). Mental Models: Towards a Cognitive Science of Language, Inference, an Consciousness. Cambridge: Cambridge University Press.
Johnson-Laird, P.N. (1989). Mental Models. In M.L. Posner, ed., Foundations of Cognitive Science. Cambridge, MA: MIT Press.
Knoblich, G., Elsner, B., von Aschersleben, G. und Metzinger, T. (2003)[Hrsg.]. Self and Action. Special issue of Consciousness & Cognition (12:4), December 2003. doi:10.1016/j.concog.2003.08.009.
Lenggenhager, B., Tadi, T., Metzinger, T. & Blanke, O.). Video Ergo Sum: manipulating bodily self-consciousness. Science, 317, 1096-9. doi:10.1126/science.1143439.
Metzinger, T. & Gallese, V. (2003). The emergence of a shared action ontology: building blocks for a theory. In G. Knoblich, B. Elsner, G. von Aschersleben, und T. Metzinger (eds), Self and Action. Special issue of Consciousness & Cognition (12:4), December 2003, 549-571. doi:10.1016/S1053-8100(03)00072-2.
Metzinger, T. (2000). Neural Correlates of Consciousness: Empirical and Conceptual Questions. Cambridge, MA: MIT Press.
Metzinger, T. (2003a). Being No One. The Self-Model Theory of Subjectivity. Cambridge, MA: MIT Press.
Metzinger, T. (2003b). Phenomenal transparency and cognitive self-reference. Phenomenology and the Cognitive Sciences, 2, 353-393. doi:10.1023/B:PHEN.0000007366.42918.eb
Metzinger, T. (2004a). Précis of „Being No One“. In

PSYCHE - An Interdisciplinary Journal of Research on Consciousness, 11 (5), 1-35. http://psyche.cs.monash.edu.au/symposia/metzinger/precis.pdf
Metzinger, T. (2004b). The subjectivity of subjective experience: A representationalist analysis of the first-person perspective. Networks, 3-4, 33-64.
Metzinger, T. (2004c). Why are identity-disorders interesting for philosophers? In Thomas Schramme und Johannes Thome (eds.), Philosophy and Psychiatry. Berlin: de Gruyter.
Metzinger, T. (2005). Out-of-body experiences as the origin of the concept of a "soul". Mind and Matter, 3(1), 57-84.
Metzinger, T. (2006). Conscious volition and mental representation: Towards a more fine-grained analysis. In N. Sebanz und W. Prinz (Hrsg.), Disorders of Volition. Cambridge, MA: MIT Press. S. 19-48.
Metzinger, T. (2008). Empirical perspectives from the self-model theory of subjectivity: A brief summary with examples. In Rahul Banerjee and Bikas K. Chakrabarti (eds.), Progress in Brain Research, 168: 215-46. Amsterdam: Elsevier. [Electronic offprint available from author]
Metzinger, T. (1995). Conscious Experience. Thorverton: Imprint Academic & Paderborn: mentis.
Metzinger, T. (2003a). Being No One. The Self-Model Theory of Subjectivity. Cambridge, MA: MIT Press. [2nd revised edition September 2004]

Metzinger, T. (2011). The no-self alternative. In S. Gallagher (ed.), The Oxford Handbook of the Self. Oxford, UK: Oxford University Press. Pp 279-296. doi:10.1093/oxfordhb/9780199548019.003.0012.
Schilling, M. & Cruse, H. (2008). The evolution of cognition - from first order to second order embodiment. In I. Wachsmuth & G. Knoblich (eds.), Modeling Communication with Robots and Virtual Humans doi:10.1007/978-3-540-79037-2_5.

(Berlin:Springer),77–108.

Schilling, M. & Cruse, H. (2012). What's Next: Recruitment of a Grounded Predictive Body Model for Planning a Robot's Actions. Frontiers in Psychology (3), 10.3389/fpsyg.2012.00383.
Seth, A.K, Suzuki, K., and Critchley, H.D. (2012). An interoceptive predictive coding model of conscious presence. Frontiers in Psychology: Consciousness Research. 2:e395 doi:10.3389/fpsyg.2011.00395.
Weisberg, J. (2005). Consciousness Constrained: A Commentary on Being No One. In PSYCHE - An Interdisciplinary Journal of Research on Consciousness, 11 (5). http://psyche.cs.monash.edu.au/symposia/metzinger/Weisberg.pdf
Windt, J. & Metzinger, T. (2007). The philosophy of dreaming and self-consciousness: What happens to the experiential subject during the dream state? In D. Barrett & P. McNamara (eds.), The New Science of Dreaming.

Volume 3: Cultural and Theoretical Perspectives. Westport, CT & London: Praeger Imprint/Greenwood Publishers. Pp. 193-247.

Internal references

Valentino Braitenberg (2007) Brain. Scholarpedia, 2(11):2918. doi:10.4249/scholarpedia.2918.
Olaf Sporns (2007) Complexity. Scholarpedia, 2(10):1623. doi:10.4249/scholarpedia.1623.
James Meiss (2007) Dynamical systems. Scholarpedia, 2(2):1629. doi:10.4249/scholarpedia.1629.
Seiji Ogawa and Yul-Wan Sung (2007) Functional magnetic resonance imaging. Scholarpedia, 2(10):3105. doi:10.4249/scholarpedia.3105.
Walter J. Freeman (2007) Intentionality. Scholarpedia, 2(2):1337. doi:10.4249/scholarpedia.1337.
Mark Aronoff (2007) Language. Scholarpedia, 2(5):3175. doi:10.4249/scholarpedia.3175.
Philip Holmes and Eric T. Shea-Brown (2006) Stability. Scholarpedia, 1(10):1838. doi:10.4249/scholarpedia.1838.
Anthony T. Barker and Ian Freeston (2007) Transcranial magnetic stimulation. Scholarpedia, 2(10):2936. doi:10.4249/scholarpedia.2936.

See Also

Consciousness

Sponsored by: Anil Seth, University of Sussex, UK

Reviewed by: Anonymous

Accepted on: 2007-10-25 06:58:04 GMT