PSYCHE

Précis: Being No One

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

This is a short sketch of some central ideas developed in my recent book *Being No One* (BNO hereafter). A more systematic summary, which focuses on short answers to a set of specific, individual questions is already contained *in* the book, namely as BNO section 8.2. Here, I deliberately and completely exclude all work related to semantically differentiating and empirically constraining the philosophical concept of a "quale" (mostly Chapter 2, 3 & 8), all proposals regarding conceptual foundations for the overall theory (2 & 5), all of the neurophenomenological case-studies used to test and refine it (4 & 7), and all remarks of a more general or methodological character (1 & 8). In particular, the present Précis does *not* follow the structure of the book. Instead, it simply sums up what the theory has to say about its three major epistemic targets, namely about consciousness (Section 2), the phenomenal self (Section 3), and the emergence of a first-person perspective (Section 4).

The Self-Model Theory of Subjectivity (SMT hereafter) develops a *constraint-satisfaction approach* to phenomenal experience. It offers a catalogue of 10 constraints for conscious representations. I have here picked 6 out of the original 10 constraints and will go through them, in order to illustrate the general approach. Before doing so, I will now (Section 1) very briefly sketch what the present theory has to say about consciousness, the phenomenal self and the first-person perspective. We will then go into slightly more detail, by describing the relevant constraints (Sections 2, 3 & 4).

1. SMT: A First Sketch

1.1. Consciousness

The present theory develops a detailed story about precisely what properties representations in a given information-processing system must possess in order to become *phenomenal* representations, ones the content of which is at the same time a content of consciousness. Let us start with what I call the "minimal concept of consciousness" and then proceed to enrich it. Phenomenologically, minimal consciousness is described as *the presence of a world*. This minimal notion involves what is called (1) the globality-constraint, (2) the presentationality-constraint, and (3) the transparency-constraint.

1.1.1. Globality Mental representation is the process by which some biosystems generate an internal depiction of parts of reality. Not all mental states are also conscious states: Phenomenally represented information is precisely that subset of currently active information in the system, of which it is true that it is **globally available** for many different processing capacities at the same time, e.g., for deliberately guided attention, cognitive reference, and the selective control of action. To say that the contents of conscious experience are "globally" available for the subject means that these contents can always be found *in a world*. This implies that individual conscious states, in standard situations, are always part of an integrated **world-model.** More about this in Section 2.1.

1.1.2 Presentationality A second core-aspect of phenomenal consciousness is what could be described as the generation of an island of presence in the continuous flow of physical time (Ruhnau 1995): Without exception, it is true of all my phenomenal states that whatever I experience, I always experience it *now*. Phenomenal content invariably is content *de nunc*, because it is associated with a representation of **temporal internality**. There is an overarching representational context governing phenomenal experience, and this context generates the **experience of presence**.

1.1.3. Transparency The third constraint for phenomenal consciousness is transparency. It is a phenomenological concept (and not an epistemological one) which, however, implies a *lack* of knowledge. Transparency is a special form of darkness. In particular, **phenomenal transparency** means that something particular is not accessible for subjective experience, namely the representational nature of the contents of conscious experience. What makes a phenomenal representation transparent is the **attentional unavailability of earlier processing stages** in the brain for introspection. The instruments of representation themselves cannot be represented as such, and hence the system making the experience, on this level and by conceptual necessity, is entangled in a **naïve realism**: In standard configurations, one's phenomenal experience has an untranscendably realistic character.

Naïve realism can also be accommodated on the epistemological level by introducing the concept of "**autoepistemic closure**". It is an epistemological, and not (at least not primarily) a phenomenological concept. It refers to an "inbuilt blind spot", a structurally anchored deficit in the capacity to gain knowledge about oneself. Specifically, the autoepistemic closure consists in human beings in ordinary waking states not being able to realize the simple fact that the content of their subjective experiences inevitably has strong, self-constructed aspects, because it is *representational* content, and that it always is simulational, counterfactual content.

1.1.4. Minimal consciousness The conjunction of satisfied constraints 1, 2 and 3 (globality, presentationality and transparency) yields the most elementary form of conscious experience conceivable: The presence of a world. The phenomenal presence of a world is the activation of a coherent, global model of reality (**Constraint 1**) within a virtual window of presence (**Constraint 2**), a model that cannot be recognized *as* a model by the system generating it within itself (**Constraint 3**). Please note how all that such a system would experience would be the presence of one unified world, homogeneous and frozen into an internal Now as it were. Neither a rich internal structure nor the complex texture of subjective time or the perspectivalness going along with a first-person point of view exists at this point. We could call this "Selfless Snapshot Consciousness".

In Section 2, I will further develop the three minimal constraints just described and three additional ones (Convolved holism, Dynamicity, Perspectivalness). This will allow us to better understand both the minimal form of consciousness and its evolution to more sophisticated forms. However, the current description of consciousness already allows us to state the position of the present theory on the phenomenal self (Section 1.2) and on the first-person perspective (Section 1.3).

1.2. The Phenomenal Self

First, it is important to understand the central ontological claim put forward by SMT: **No such things as selves exist in the world**. For all scientific and philosophical purposes, the notion of a self – as a theoretical entity – can be safely eliminated. What we have been calling "the" *self* in the past is not a substance, an unchangeable essence or a thing (i.e., an "individual" in the sense of philosophical metaphysics), but a very special kind of representational content: The content of a self-model that cannot be recognized *as* a model by the system using it. The dynamic content of the **phenomenal *self*-model** (hereafter: ”**PSM**”, cf. BNO, Chapter 6) is the content of the conscious self: Your current bodily sensations, your present emotional situation plus all the contents of your phenomenally experienced cognitive processing. They are constituents of your PSM. All those properties of your experiential self, to which you can now direct your attention, form the content of your current PSM. This PSM is not a thing, but an integrated process.

Intuitively, and in a certain metaphorical sense, one could say that you *are* the content of your PSM. A perhaps better way of making the central point intuitively accessible could be by saying that we are systems that constantly *confuse* themselves with the content of their PSM. At least for all conscious beings so far known to us it is true that they neither *have* nor *are* a self. Biological organisms exist, but an organism is not a self. Some organisms possess conscious self-models, but such self-models certainly are not *selves* – they are only complex brain states. However, if an organism operates under a transparent self-model, then it possesses a *phenomenal* self. The phenomenal property of selfhood as such is a representational construct: an internal and dynamic representation of the organism as a whole to which the transparency constraint applies. It truly is a *phenomenal* property in terms of being an appearance only. The phenomenal experience of **substantiality** (i.e., of being an independent entity that could in principle exist all by itself), of having an **essence** (i.e., of being defined by possessing an unchangeable innermost core, an invariant set of intrinsic properties) and of **individuality** (i.e., of being an entity that is unique and indivisible) are special forms of conscious, representational

content as well. Possessing this content on the level of phenomenal experience was evolutionary advantageous, but as such (i.e., as *phenomenal* content) it is not epistemically justified.

This position is clearly counter-intuitive: For human beings, during the ongoing process of conscious experience characterizing their waking and dreaming life, *a self is present*. Human beings consciously experience themselves as *being someone*. The problem for the present theory thus is to explain how one's own **personal identity** appears in conscious experience: What is needed to—by conceptual necessity—take the step from the representational property of self-modeling to the consciously experienced phenomenal property of selfhood?

My claim is that the transparency-constraint (as defined above, for more see Section 2.3 and 3.3) is the decisive defining characteristic: If all other necessary/sufficient constraints for the emergence of phenomenal experience are satisfied by a given representational system, the addition of a transparent self-model will by necessity lead to the emergence of a phenomenal self. The transparency of the self-model is a special form of inner darkness. It consists in the fact that the representational character of the contents of self-consciousness is not accessible to subjective experience. The phenomenology of transparent self-modeling is the phenomenology of selfhood. It is the phenomenology of a system caught in a **naive-realistic self-misunderstanding**. A selfless system can certainly misunderstand itself, for instance by misinterpreting phenomenal experience in terms of implying the actual existence of a self. Phenomenal selfhood results from autoepistemic closure in a self-representing system; it is a function realized by a *lack* of information. We do not experience the contents of our self-consciousness as the contents of a representational process, but simply as *ourselves, living in the world right now*.

1.3. The Consciously Experienced First-Person Perspective

The existence of a coherent self-representatum[1] for the first time introduces a **self-world-border** into the system's model of reality. For the first time, system-related information now becomes globally available *as system-related* information, because the organism now has an internal image of itself as a whole, as a distinct entity possessing global features. On the other hand, environment related information can now be referred to as *non-self*. Objectivity emerges together with subjectivity. The functional relevance of this way of generating a very fundamental partitioning of representational content into two very general classes lies in the way in which it forms a necessary pre-condition for the activation of more complex forms of phenomenal content: *Relations* between the organism and varying objects in its environments can now for the first time be consciously represented. A system that does not possess a coherent, stable self-representatum is unable to internally represent all those aspects of reality associated with self-world, self-object, and, importantly, self-self and self-other relations. Let us call this the "**principle of phenomenal intentionality-modeling**": complex information pertaining to dynamical subject-object relations can be extracted from reality and used for selective and flexible further processing only, if a conscious self-model is in existence. Moreover, a deeper (but frequently neglected) phenomenological truth about perception can be rediscovered on higher levels of self-consciousness: Full-blown, phenomenal self-

consciousness always involves a relation between the self and an object-component. The content of a perceptual state really is *not* a part of the environment, but a *relation* holding to this part. This is also true in inner environments.

The ongoing, episodic subject–object-relation is the content of what I call the "**phenomenal model of the intentionality relation**" (**PMIR** hereafter, for further details see Section 4; BNO Chapter 6, Metzinger 2005; for some ideas about its unconscious functional precursors in the monkey brain see Metzinger & Gallese 2003). Here are four different examples, in terms of typical phenomenological descriptions of the class of phenomenal states at issue: ''I am someone, who is currently visually attending to the colour of the book in my hands,'' ''I am someone currently grasping the content of the sentence I am reading,'' ''I am someone currently hearing the sound of the refrigerator behind me,'' ''I am someone now deciding to get up and get some more juice.''

The central defining characteristic of phenomenal models of the intentionality-relation is that they depict a certain *relationship* as currently holding between the system as a whole, as transparently represented to itself, and an object-component. It is easy to see how for the first time PMIR allows a system to consciously experience itself as being not only a part of the world, but of being fully immersed in it through a dense network of causal, perceptual, cognitive, attentional, and agentive relations. The core-idea behind the notion of a PMIR is that the decisive feature characterizing the representational architecture of human consciousness lies in **continuously co-representing the representational relation itself**.

1.4. A Closer Look at SMT

In the remaining sections, I will detail 6 relevant constraints for the definition of consciousness, the phenomenal self and the first-person perspective. In BNO, these constraints are developed on the four most important levels of analysis: the *phenomenological* level of description (which operates from the first-person perspective, trying to give fine-grained and clear descriptions of the target phenomena), the *representationalist* level (which analyses these targets as forms of representational content), the *functionalist* level (describing causal roles and computational features), and the *neurobiological* level (which, wherever this is already possible, points to potential neural correlates in the domain of biological systems). Here, I will sometimes have to skip the corresponding sections in the interest of brevity.

2. Six Constraints for Consciousness

2.1. Globality

The globality constraint is a differentiated and refined version of the notion of "global availability"(see Baars, 1988, 1997; Chalmers 1997). At least one important limitation to this principle is known. A large majority of simple sensory contents is not available for cognitive reference, because **perceptual memory** cannot grasp contents that are internally individuated in a fine-grained manner. In general, however, all phenomenal representata make their content at least globally available for attention and motor control, even if *not* for mental concept formation. A first consequence of this is that for

possessing phenomenal experience it is not necessary to at the same time possess capacities for mental concept formation or linguistic abilities.

The globality constraint demands that individual phenomenal events are always bound into a **global situational context** (for a potential exception, see the discussion of akinetic mutism in BNO 6 & 8, plus Metzinger 2005). In the terms used in the present theory, it means that individual conscious states, in standard situations, are always part of a conscious **world-model**. We can transport this constraint from the subpersonal to the personal level of description, by making the following statement: If and only if a person is conscious, a world exists for her, and if and only if she is conscious, she can make the fact of actually living *in* a world available for herself, cognitively and as an agent.

2.1.1. The phenomenology of global availability The contents of conscious experience are characterized by my ability to react directly to them with a multitude of my mental and bodily capacities: I can direct my attention toward a perceived color or toward a bodily sensation, in order to inspect them more closely ("attentional availability"). In at least some cases I am able to form thoughts about this particular color. I can make an attempt to form a categorical representation, a concept for it ("availability for phenomenal cognition"), which associates it with earlier color experiences ("availability for autobiographical memory") and I can communicate about color with other people by using language ("availability for speech control", which might also be termed "communicative availability"). I can reach for colored objects and sort them according to their phenomenal properties ("availability for the control of action"). In short, global availability is an all-pervasive functional property of my conscious contents, which *itself* I subjectively experience – namely as my own flexibility and autonomy in dealing with these contents.

2.1.2. Globality as a representational property Phenomenal representata are characterized by the fact of their intentional *content* being directly available for further processing by subsymbolic mechanisms like attention or implicit memory, but also for concept formation, for metacognition and verbal report, for planning or for motor simulations with immediate behavioral consequences. Its globality consists in being embedded into a functionally active model of the world at any point in time (Yates 1985), into an overarching, singular and coherent representation of reality as a whole. Specifically, three aspects of this world-model are of particular interest on the representational level of description: the **numerical identity** of the reality depicted by it, its **coherence** and the constant **dynamical integration** of individual contents leading to this coherence.

2.1.3. Globality on the functional level of analysis: The generation of an inner world as an informational/computational strategy In the brain, there is no such thing as a truly "final" phase of processing. But the generation of a single and coherent world-model is a strategy to achieve a **reduction of ambiguity** originating in the buzzing, blooming confusion of the external world. At the same time, this leads to a reduction of *data*: the amount of information directly available to the system, e.g., for selection of motor processes or the deliberate guiding of attention, is being minimized and thereby, for all mechanisms operating on the phenomenal world-model, computational load is reduced.

The functionalist reading of our first constraint is this: phenomenal representata, by conceptual necessity, are operated on by a highest-order integrational function. This function dynamically unifies a large number of microcausal relations into a single and **distinct causal role**. Elsewhere, I have introduced a speculative concept, the concept of *highest-order binding*, in short "HOB" (Metzinger 1995b). Representational states, after being integrated into the phenomenal world-model, can interact with a very large number of specialized modules in very short periods of time and in a context-sensitive, flexible manner, thereby also increasing the adaptive flexibility of the system's behavioral profile. The more information is conscious, the higher the degree of flexibility and context-sensitivity of its reactions to the environment will be, because many different functional modules can now access and use this information in a direct manner to react to challenges from the environment in a differentiated way. One-step updating/learning becomes possible. But new cognitive dimensions open up as well: Only if you have the subjective experience of a world being present right now can you start to conceive of the notion of a single reality. In our own case, even the appearance-reality distinction becomes attentionally as well as cognitively available.

2.1.4. Neural correlates of global integrational functions Currently no detailed theories concerning the possible neural correlates, in particular of the minimally *sufficient* correlate (Chalmers 2000), for the appearance of a coherent, conscious model of the world do exist. However, there are a number of interesting speculative hypotheses.

A first core intuition has been to study the mechanism of action common to different anaesthetics, that is, to study the conditions under which phenomenal experience *as a whole* disappears and re-emerges (for further references and a recent discussion of the potential role of the NMDA-receptor-complex in achieving large-scale integrations of ongoing activity, see Flohr 2000, Franks and Lieb 2000, Hardcastle 2000, and Andrade 2000).

A second important insight is that the globality-constraint applies to two fundamentally different classes of phenomenal states: to dreams (see BNO, Section 4.2.5) and to waking states. In dreams, as well as during ordinary waking phases, the system operates under one single, more or less coherent world-model while its global functional properties differ greatly. Rodolfo Llinás and his co-workers have long emphasized that one of the most fruitful strategies in searching for the NCC will be in "subtracting" certain global properties of the waking world-model from the dreaming world-model, thereby arriving at a common neurophysiological denominator or at global functional states which are basically equivalent between phenomenal experience during REM-sleep and waking (Llinás and Paré 1991, p. 522 pp.). The intuition behind this neuroscientific research program carries a distinct philosophical flavor: What we call waking life is a form of "online dreaming". If there is a common functional core to both global state-classes, then conscious waking would then be just a dreamlike state that is currently modulated by the constraints produced by specific sensory input (Llinás and Ribary 1993, 1994; Llinás and Paré 1991). A specific candidate for a global integrational function offered by Llinás and colleagues is a rostrocaudal 12-ms phase shift of 40-Hz-activity related to synchronous activity in the thalamocortical system, modulated by the brain stem (the most detailed presentation of Llinás' thalamocortical model can be found in

Llinás and Paré 1991, p. 531; see also Llinás and Ribary 1992; Llinás, Ribary, Joliot and Wang 1994; Llinás and Ribary 1998; Llinás, Ribary, Contreras and Pedroarena 1998).

The strategy of approaching the globality-constraint by researching globally coherent states (as initially proposed in Metzinger 1995b) leads to a new way of defining research targets in computational neuroscience (e.g., von der Malsburg 1997). However, it must be noted that what is actually needed is a theoretical model that allows us to find global neural properties exhibiting a high degree of integration and differentiation *at the same time*. The neural correlate of the global, conscious model of the world must be a distributed process which can be described as the realization of a *functional cluster*, combining a high internal correlation strength between its elements with the existence of distinct functional borders. This cluster directly corresponds to the distinct causal role mentioned above. Edelman and Tononi have called this the "dynamic core hypothesis" (see Tononi and Edelman 1998a,b; Edelman and Tononi 2000a, Tononi 2003; for a comprehensive popular account see Edelman and Tononi 2000b). The hypothesis states that any group of neurons can contribute directly to conscious experience only if it is part of a distributed functional cluster that, through reentrant interactions in the thalamocortical system, achieves high integration in hundreds of milliseconds. At the same time it is essential that this functional cluster possesses high values of complexity. Converging evidence seems to point towards a picture in which large-scale integration is mediated by the transient formation of dynamical links through neural synchrony over multiple frequency bands (Varela, Lachaux, Rodriguez, and Martinerie 2001, Singer 2004).

This way of looking at the globality-constraint on the neural level is philosophically interesting for a number of reasons. First, it makes the prediction that any system operating under a conscious model of reality will be characterized by the existence of one *single* area of **maximal causal density** within its information-processing mechanisms. To have an integrated, globally coherent model of the world means to create a global functional cluster, i.e., an island of maximal causal density within one's own representational system. Philosophical functionalists will like this approach, because it offers a specific and global functional property (a "vehicle property") that might correspond to the global phenomenal property of the **unity of consciousness**. In short, what you subjectively experience upon experiencing your world as coherent is the high internal correlation strength between a subset of physical events in your own brain. Second, it is interesting to note how the large group of neurons constituting the dynamical core in the brain of an organism currently enjoying an integrated conscious model of reality will very likely be different at every single instant. The physical composition of the core state will change from millisecond to millisecond. At any given point in time there will be one global, minimally sufficient neural correlate of consciousness, but at the next instant this correlate will already have changed, because the consciousness cluster only constitutes a *functional* border which can easily transgress anatomical boundaries from moment to moment. Third, it has to be noted that the informational content of the dynamical core is determined to a much higher degree by internal information already active in the system than by external stimuli. Just as in the Llinás-model, an overall picture emerges of the conscious model of reality essentially being an internal construct, which is only *perturbed* by external events forcing it to settle into ever-new stable states.

In short, there may be many functional bundles - individual and episodically indivisible, integrated neural processes - within a system, and typically there will be *one* single, largest island of maximal causal density underlying the current conscious model of the world. "Indivisible" here means that although one could, from a third-person perspective, still find causal divisions and a fine-structure characterizing the events causally integrated by such a process. *For a certain period of time*, they would be indivisible for the system, in which they occur, because the sheer dynamical coherence of these bundles would be something which the system, using its own resources of causal interaction, could not dissolve. Still, the philosophical question remains of what it is that makes this cluster into the *subjective* world the organism lives in. It is plausible to assume that at any given time this typically is the largest functional cluster (for a dissenting view, see Zeki and Bartels 1998). Nevertheless, the question remains how such a cluster becomes tied to an individual first-person perspective, to a representation of the system itself, and thereby becomes a truly *subjective* global model of reality (see the perspectivalness-constraint in BNO, Section 3.2.6 and Chapter 6; plus Section 4 below).

2.2. Presentationality

The world that appears to me does so by *being present*. The experience of presence coming with our phenomenal model of reality may be the central aspect that cannot be "bracketed" in a Husserlian sense: It, as it were, is the temporal immediacy of existence *as such*. If we subtract the global characteristic of presence from the phenomenal world-model, then we simply subtract its existence. We would subtract consciousness *tout court*. It would not *appear* to us anymore. Let us take a closer look.

2.2.1. The phenomenology of presence To consciously experience means to *be in a present*. It means that you are processing information in repeatedly and continuously integrating individual events (already represented as such) into larger temporal *Gestalts*, into one singular psychological moment. What is a conscious *moment*? The phenomenal experience of time in general is constituted by a series of important achievements. They consist in the phenomenal representation of temporal identity (experienced simultaneity), of temporal difference (experienced non-simultaneity), of seriality and unidirectionality (experienced succession of events), of temporal wholeness (the generation of a unified present, the "specious" phenomenal Now) and the representation of temporal permanence (the experience of duration). The decisive transition toward subjective experience, that is, toward a genuinely *phenomenal* representation of time takes place in the last but one step: Precisely when event representations continuously are integrated into psychological *moments*.

The feature that is conceptually so hard to grasp is how we can consciously experience a full-blown present *as embedded into* a uni-directional flow, the experience of duration. There are temporal *Gestalts*, islands of individually characterized Nows, but the background against which these islands are segregated is itself nothing static: it possesses a direction.

2.2.2. Phenomenal presence as a form of representational content *de nunc* Proceeding to the representationalist level of description we first find that the phenomenal processes of representation not only generate spatial, but also *temporal internality*: There is a specific *de-nunc*-character of phenomenal content. A crucial point

to note when shifting back into the third-person perspective is that the physical world is "nowless", as well as it is futureless and pastless. A complete physical description of the universe would not contain any information about what time is "now", nor an analysis of time as a unidirectional phenomenon. On the contrary, the conscious experience of time inevitably possesses an **indexical component in the temporal domain**. This type of mental content is *simulational*: It is not an epistemically justified form of content in that, strictly speaking, it does not involve knowledge about the current state of the actual world. Although we subjectively experience ourselves as in direct and immediate contact with the "Now", all empirical data tell us that, strictly speaking, all conscious experience is a form of memory. Information represented by phenomenal models of reality is always being presented to the subject of experience as *actual* information. It is this form of temporal internality which is a simulational fiction from the third-person perspective.

2.2.3. The window of presence as a functional property Generally speaking, any purely data-driven model of the world will not permit explicit predictions in time (Cruse 1999). Only additional, recurrent networks will allow for the generation of time-dependent states. The representation of a "Now" then becomes the *simplest* form of explicit time representation, as a set of recurrent loops plus a certain decay function. Obviously, **short-term and working memory** will be at the heart of any cognitivist/functionalist analysis of the presentationality-constraint for phenomenal content.

2.2.4. Neural correlates of the window of presence Very little is known in terms of implementational details. Ernst Pöppel, in a whole series of publications, has emphasized how certain empirically well-documented oscillatory phenomena in the brain could serve as providing a rigid internal rhythm for internal information-processing, namely by generating "elementary integration units" (EIUs; this is Pöppel's terminology, see Pöppel 1994, 1995, see also 1988). The generation of such EIU can be interpreted as a process of internal data-reduction: the system *deletes* information about its own physical processuality, by not defining temporal relations between elements given within such a basal window of simultaneity. Using philosophical terminology, we might say that the physical temporality of the actual *vehicles* participating in this elementary representational process, thereby, is not reflected on the level of their *content* anymore. The fine structure of physical time is now internally invisible for the system, by becoming *transparent* (see BNO, section 3.2.7 and Section 2.3 below).

2.3 Transparency

Phenomenal transparency certainly is one of the (if not *the*) most important constraints, if we want to achieve a deeper theoretical understanding of what phenomenal experience really is. Therefore, we have to avoid any confusion with existing notions of "transparency" (See BNO, section 3.2.6 and Metzinger 2003b for more).

Today, 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[2], 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 – in particular, its overgeneralized use - is unsatisfactory, because it violates important phenomenological constraints: Introspective unavailability of the carrier 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.

Let me thus introduce my own working definition of phenomenal transparency: 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**. Under SMT, we are exclusively concerned with **phenomenal transparency**, thus unconscious representations are neither transparent nor opaque. That is to say, transparency is a property of active mental representations already satisfying the minimally sufficient constraints for conscious experience to occur. Notably, phenomenally transparent representations are always activated within a virtual window of presence and are functionally integrated into a unified global model of the world.

In this context, the following fact is of particular philosophical interest: The more earlier processing stages, and the more earlier aspects of the internal construction process leading to the final, explicit and disambiguated phenomenal content, are available for introspective attention, the more will the system be able to recognize these phenomenal states *as* internal, self-generated constructs. Full transparency means full attentional unavailability of earlier processing stages. Degrees of opacity come as degrees of attentional availability. Hence, the following principle: For every phenomenal state, the **degree of phenomenal transparency is inversely proportional to the introspective degree of attentional availability of earlier processing stages**.

2.3.1 The phenomenology of transparency What is inaccessible to conscious experience is the simple fact of this experience taking place in a medium. Therefore, transparency of phenomenal content leads to a further characteristic of conscious experience, namely the subjective impression of **immediacy**. Many bad philosophical arguments concerning direct acquaintance, infallible first-person knowledge and direct reference are based on an equivocation between epistemic and phenomenal immediacy: from the fact that the conscious experience, e.g., of the color of an object, carries the characteristics of phenomenal immediacy and direct givenness it does not follow that any kind of non-mediated or direct kind of knowledge is involved.

Many authors describe phenomenal transparency as an all-or-nothing phenomenon. To do phenomenological justice to conscious experience, however, demands a more differentiated description. Sensory experience is the paradigmatic example of fully transparent phenomenal content. There are, however, examples of sensory opacity, for instance, during extremely short transition phases in bistable phenomena, e.g., if a consciously experienced Necker cube switches from one interpretation to the next and back, or during the phenomenon of binocular rivalry (see, e.g., Leopold and Logothetis 1999). In conscious thought we have the paradigm example for opacity, with the manifest daydream being the exception. Emotions lie in between, and exhibit much greater variability. These simple phenomenological observations point to an important functional characteristic of opaque phenomenal representations: They make the possibility that they

actually might be **misrepresentations** globally available for cognition, attention and behavioral control.

When discussing the phenomenology of transparency and opacity, it must be noted that not only can individual phenomenal contents exhibit a variable degree of transparency; the same is true of global phenomenal world-models as well. Right after a traffic accident the whole world can appear as "unreal" or as "dreamlike" to us. The same phenomenon is known in stress situations and in transitory phases during certain psychiatric syndromes ("derealization"). However, the best and most basic example for an almost fully opaque, global phenomenal state is the lucid dream (see LaBerge and Gackenbach 2000 BNO section 7.2.5).

At this point, one of the philosophically most interesting issues lies in the question if transparency *really* is a necessary condition for phenomenality. If it is, how then do we account for the cases of opaque phenomenal representations just mentioned? Moreover, is the account here sketched not trivially circular, simply because the concept of transparency is already introduced as a property of phenomenal representations only? I can only give an answer in Section 3.3, because only then we can understand why phenomenal transparency is important in understanding the *subjectivity* of our target phenomenon. For now, let us note that the lucid dream could appear as a candidate of a global conscious state in which *everything* is experienced as the content of a representation in one's own mind, thus satisfying constraint 1 & 2 but not constraint 3. However, upon closer inspection we see that the consciously represented **subject of experience**, the lucid dreamer's phenomenal self does not itself appear as representational content – it still looks entirely real. There still is *someone having* the dream. The interim conclusion is that a necessary minimal degree of transparency may exist for any form of conscious experience. This phenomenological point and its conceptual consequences is what we must return to below (see section 3.3).

Before moving on to the representational level of description, please note how there exist three important **equivocations** or potential misunderstandings of the notion of "phenomenal transparency," as introduced here (see Metzinger 2003b for details). First, transparency is not an epistemological notion, but a phenomenological concept. In particular, it has nothing to do with the Cartesian notion of *epistemic transparency*, the philosophical intuition that in principle I cannot be wrong about the content of my own consciousness, that the notion of an unnoticed error in introspectively accessing the content of your own mind is incoherent. Moreover, transparency is here conceived of as a property of phenomenal representations in a sub-symbolic medium, i.e., of non-linguistic entities under an empirically plausible theory of mental representation, and not as a property of a context. The second potential equivocation is the extensionality-equivocation: Transparency as a property of extensional (i.e., *referentially* transparent) contexts is something entirely different. Phenomenal transparency can exist in non-linguistic creatures, lacking any form of cognitive reference. There is a third common use of the notion "transparency," which should not be confused with the notion as here intended: In communication theory it is conceived of as a property of *media*. For instance, in technical systems of telecommunication, transparency can be the property of a channel or of a system for the transmission of information in general.

2.3.2 Transparency as a property of conscious representations Phenomenal representations are transparent, because their content appears to be fixed in all possible contexts: The paper, which you now hold in your hands, will always stay this very paper according to your subjective experience, no matter how much the external perceptual situation changes. At this level it may be helpful to clarify the concept of transparency with regard to the current theoretical context by returning to more traditional conceptual tools, by differentiating between the vehicle and the content of a representation.

The representational carrier of your phenomenal experience is a certain process in the brain. This process, which in no concrete way possesses anything "paper-like" - is not consciously experienced by yourself, it is transparent in the sense of you looking through it. What you are looking *onto* is its representational content, the existence of a paper, here and now, as given through your sensory organs. This content, therefore, is an abstract property of the concrete representational state in your brain. However, there are at least two kinds of content. The **intentional content** of the relevant states in your head in its epistemic status depends on the fact of this paper actually existing, and of the relevant state being a reliable instrument for gaining knowledge in general. If this representational carrier is a good and reliably functioning instrument for generating knowledge about the external world, then, by its very transparency, it permits you to directly, as it were, look "through it" right onto the paper. It makes the information carried by it globally available (Constraint 1), without you having to care about *how* this little miracle is achieved. The **phenomenal content** of your currently active paper representation is what stays equal, no matter if the paper exists or not. It is solely determined by internal properties of the nervous system. If your current perception, unnoticed by you, actually is a hallucination, then, as it were, you, as the system as a whole, are not anymore looking "through" the state in your head onto the world, but only at the representational vehicle itself - without *this* fact itself being globally available to you. The specific and highly interesting characteristic of the phenomenal variant of representation now is the fact that this content, even in the situation just described, is invariably experienced as maximally *concrete*, as absolutely unequivocal, as maximally determinate and disambiguated, as directly and immediately given to you.

2.3.3. Transparency as an informational/computational strategy Transparency of internal data-structures is a great advantage for any biosystem having to operate with limited temporal and neural resources. Indeed, it minimizes computational load since it is synonymous to a *missing* of information on this level of processing: Our representational architecture only allows for a very limited introspective access to the real dynamics of the myriads of individual neural events out of which our phenomenal world finally emerges in a seemingly effortless manner.

Naïve realism hinders the system to lose contact with external reality by getting lost in an introspective exploration of the underlying mechanisms. On the epistemological level, naïve realism corresponds to what I call "autoepistemic closure". Autoepistemic closure as used in BNO does not refer to cognitive closure (McGinn 1989, 1991) or epistemic "boundedness" (Fodor 1983) in terms of the unavailability of theoretical, propositionally structured self-knowledge. Rather, it refers to a closure or boundedness of *attentional processing* with regard to one's own, internal representational dynamics:

Subjective experience has not been developed in pursuing the old philosophical ideal of self-knowledge.

2.3.4 Transparency as a functional property Systems operating under a transparent world-model for the first time live in a reality, which, for them, cannot be transcended: On a functional level they become *realists*. Again, this does not mean that they have to possess or even be able to form certain beliefs, or use explicit symbol structures in communication. It means that the implicit assumption of the actual presence of a world becomes causally effective. The transparent world-model allows a system to treat information as *factual* information, i.e., it enables the internal **representation of facticity**. Moreover, as soon as a certain degree of opacity becomes available, a second major functional advantage emerges: The **appearance-reality distinction** can now be represented, it becomes itself an element of reality. The fact that some elements of the ongoing flow of conscious experience actually are representational *contents*, and, therefore, may be false, becomes globally available – and it is hard to underestimate the functional potential of this step.

2.3.5. The next step: Differentiated consciousness Recall how the conjunction of constraints **1**, **2**, and **3** yields the most elementary form of conscious experience: The presence of a world. The phenomenal presence of a world is the activation of a coherent, global model of reality (**Constraint 1**) within a virtual window of presence (**Constraint 2**), both of which are transparent in that the world-model cannot be recognized *as* a model by the system generating it within itself (**Constraint 3**).

Our minimal notion of consciousness is not yet subjective experience in terms of being tied to a consciously experienced **first-person perspective**. It is only subjective in the very weak sense of being an internal model within an individual organism, therefore this minimal notion still is very simplistic (and probably empirically empty), because it is completely undifferentiated in its representation of causality, space and time. A system enjoying minimal consciousness as exclusively described by the conjunction of the first 3 constraints would be frozen in an eternal Now, and the world appearing to this organism would be devoid of all internal structure.

I will now (more briefly) describe three additional constraints allowing for the characterization of more sophisticated forms of consciousness. If we add a mereological internal structure in terms of **Constraint 4 (convolved holism)**, we allow for scene-segmentation and the emergence of a complex situation. A nested hierarchy of contents now comes into existence. However, if we do not want to assume the unlikely case of "snapshot-consciousness", of one single, presegmented scene being frozen into an eternal Now on the phenomenal level, we have to add temporal structure in terms of **Constraint 5 (dynamicity)**. At this stage it is possible to have phenomenal experience as a dynamically evolving phenomenon on the level of content, to have an interrelated hierarchy of different contents that unfolds over time and possesses a dynamical structure. **Constraint 6** is **Perspectivalness**: Consciousness is the appearance of a world from a first-person perspective.

2.4 Convolved Holism

Nestedness (or "convolution" hereafter) is a property of any hierarchical system having entities of smaller scale enclosed within those of larger scale (Salthe 1985, p. 61). Conscious experience itself can be described as a phenomenon possessing such a hierarchical structure, for instance by being composed of representational, functional and neurobiological entities assignable to a hierarchy of levels of organization.

 2.4.1. The phenomenology of embedded wholes Let us look at the paradigmatic examples for **phenomenal holism**. The lowest level on which we find an integration of features into a representational unit possessing global features like holism is the level of *perceptual* object-formation. Consciously perceived, attentionally available objects are sensory wholes, even if they are not yet linked to conceptual or memory structures. A second paradigmatic example for a holistic, coherent form of content is the phenomenal self. In standard situations, the consciously experienced self not only forms a unity, but an integrated whole. Third levels on which we find the phenomenal property of holism are complex scenes and situations: integrated arrays of objects, including relations between these objects and implicit contextual information. A visually perceived, pre-segmented scene - like a beautiful landscape you are looking at - or a complex, *multimodal* scene including sounds, smells, and a certain social context are further examples of phenomenal holism. The brief integrations between subject and object as consciously represented, the phenomenal experience of a "self in the act of knowing" is yet another paradigmatic phenomenological example for a briefly emerging integrated whole.

 What precisely does it mean to speak of "wholeness"? Holism means that, on a conceptual level, we are not able to adequately describe those aspects of a unit of experience, which can be subjectively discriminated, as isolated elements within a set. This fact is an important conceptual constraint for any serious neurophenomenology. If one only analyzes such subregions or discriminable aspects in the flow of phenomenal experience as individual components of a class, one misses one of the most essential characteristics of conscious experience. There are no decontextualized atoms. The relationship between those aspects or subregions is a **mereological relationship**. On lower levels of phenomenal granularity different aspects may be bound into different low-level wholes (different colors or smells may belong to different perceptual objects), but ultimately all of them are parts of one and the same global whole.

 2.4.2. Convolved holism as a representational property and as an informational/computational strategy Information displayed in a holistic format is highly coherent information. Phenomenal information therefore, is that subset of active information, which is available to the system in an *integrated* form. In addition, information displayed within a nested, holistic world-model generates a strong interdependence: individual property features, perceptual objects or global aspects of a scene mutually influence each other and in this way the complex causal structure of the external world can be represented with a high degree of precision. One of the functional advantages is that the representational content of a global world-model, as everything contained in it is simultaneously affecting everything else, can, in principle, be *updated* in one single step. If necessary, **local changes can effect global transitions**.

 2.4.3. Neural correlates of convolved holism Once again, we have to admit that not enough empirical data are currently available, in order to be able to make any precise

statements (but see Singer 2000, 2004, 2005; Varela, Lachaux, Rodriguez and Martinerie 2001).

In an earlier publication (Metzinger 1995b), I proposed the necessity for a subsymbolic and global integrational function that fulfils two conditions. First, that function would have to achieve global integration of representational contents active in the brain without causing a "superposition catastrophe", i.e., without causing interferences, misassociations and the mutual deletion of different representational patterns. Let us assume for a moment that the correct neurobiological theory describing the mechanism of integration, would explain it in terms of the temporal coherence of neural responses established through synchronous firing. Then the situation to be avoided would correspond to states of global synchrony as in epilepsy or deep sleep. In these states all conscious experience is typically absent. Therefore, what is needed is a function achieving a dynamical and global form of metarepresentation *by functional integration*, not simply deleting or "glossing over" all the lower-order contents, but preserving its differentiated structure. Second, the holism-producing mechanism should be conceivable as selectively operating at different levels of granularity. Therefore, what is needed to establish a differentiated type of large-scale coherence on the level of the brain itself will not be uniform synchrony, but dynamic, specific cross-system relations binding subsets of signals in different modalities and using different frequency bands at the same time (see Engel and Singer 2000; Singer 2004, 2005).

2.5. Dynamicity

In a certain sense what has just been described as convolved holism, also reappears in the phenomenology of time-experience: Our conscious life emerges from integrated psychological moments, which, however, are themselves integrated into the flow of subjective time. **Constraint 5**, **dynamicity**, does justice to the fact that phenomenal states only rarely carry static or highly invariant forms of mental content, and that they do not result from a passive, non-recursive representational process. The notion of "convolved holism" was a natural extension of the first constraint, the globality-constraint, namely on subglobal levels of description. The fifth, the dynamicity-constraint is a natural extension of the second constraint, the presentationality-constraint.

2.5.1 Phenomenology of dynamicity   The most important forms of temporal content are **presence** (as already required by constraint **2**), **duration**, and **change**. Time flows. However, the experience of flow, of duration and change is seamlessly integrated into the temporal background of presence - *all the time*, as it were. What is particularly hard to describe is the strong degree of integration holding between the experience of presence and the continuous conscious representation of change and duration. It is not as if the Now would be an island emerging in a river, in the continuous flow of consciously experienced events as it were - in a strange way the island is a *part* of the river itself.

2.6. Perspectivalness

BNO proposes a "Self-model Theory of *Subjectivity*" and subjectivity, viewed as a phenomenon located on the level of phenomenal experience, can only be understood if we find comprehensive theoretical answers to the following two questions.

First, what is a consciously experienced, phenomenal *self*? Second, what is a consciously experienced phenomenal *first-person perspective*? The remaining two sections of this Précis will entirely focus on this central aspect of the problem. However, as "perspectivalness" is also one of the centrally relevant constraints satisfied by most states of *consciousness*, I will offer a brief description of this last constraint here. It is also intended as an introduction to the remaining parts of this short Précis.

It has to be noted that perspectivalness is not a necessary condition for the ascription of conscious experience to a given system. There are a number of phenomenal state-classes - for instance, spiritual and religious experiences of a certain kind or fully depersonalized states during severe psychiatric disorders - in which an inference to the most plausible phenomenological explanation tells us that no conscious self and no consciously experienced first-person perspective exist. I take such global experiential states to be instances of *non-subjective* consciousness. On the level of their phenomenal content, they are not tied to an individual, consciously experienced first-person perspective any more. This does not mean that under a non-phenomenological, e.g., an *epistemological*, concept of subjectivity they could not still be truthfully described as weakly subjective states, for instance in terms of being exclusively internal models of reality generated by individual systems. It would be conceptually possible to describe such states as epistemically subjective, and therefore as a form of knowledge, while at the same time being phenomenally *non-subjective* in terms of not being tied to a consciously experienced first-person perspective.

2.6.1. The phenomenology of perspectivalness Perspectivalness is a **structural feature of phenomenal space** as a whole. It consists in the existence of a single, coherent, and temporally stable model of reality that is representationally centred on a single, coherent, and temporally extended phenomenal subject (Metzinger 1993, 2000b). The experiential perspectivalness of one's own consciousness is constituted by the fact that phenomenal space is centred by a phenomenal self: It possesses a focus of experience, a *point* of view. There seems to be a **primitive and prereflexive form of phenomenal self-consciousness** underlying all higher-order and conceptually mediated forms of self-consciousness (see BNO sections 5.4 and 6.4), and this non-conceptual form of selfhood constitutes the *origin* of the first-person perspective.

2.6.2. Centeredness as a functional property The experiential centeredness of our conscious model of reality has its mirror image in the centeredness of the *behavioral* space. This functional constraint is so general and obvious that it is frequently ignored: in human beings, and in all conscious systems we currently know, sensory and motor systems are *physically* integrated within the body of a single organism. This can be called the **"single embodiment-constraint"**. Consciousness not only locally supervenes, in the target domain it locally supervenes on parts of single organisms.

2.6.3. Neural correlates of the centeredness of representational space. There are many empirical results pointing to mechanisms constituting a **persisting functional link** between certain *localized* brain processes and the center of representational space. These mechanisms, for instance, include the activity of the vestibular organ, the spatial "matrix" of the body schema, visceral forms of self-representation and, in particular, the input of a number of specific nuclei in the upper brain stem, engaged in the homeostatic regulation of the "internal milieu" (see Parvizi and Damasio 2001, Damasio 1999,

Chapter 8; Damasio 2000). The function of these mechanisms consists in generating a high degree of invariance and stability, by providing the system with a continuous internal source of input. This source of input is what *anchors* the human self-model: The conscious self-model characteristically differs from all other phenomenal representations by being causally locked to this persistent functional link.

3. The PSM: Multi-level constraints for self-consciousness. What turns a neural system-model into a phenomenal self?

We just described the six constraints on an adequate theory of consciousness. These constraints can now be applied to the notion of phenomenal self-model (PSM).

Let us first recall how the SMT defines the self-model. First, from a strictly formal point of view, there exists a proof that every regulator of a complex system will automatically and by necessity become a *model* of that system (Conant & Ashby 1970). From a logical and an epistemological perspective it is helpful to differentiate between simulation and emulation, in order to further enrich the concept of a PSM. We can then, in a second step, conceptually analyze the PSM as a special variant, namely a combination of self-simulation and self-emulation. What is simulation, and what is emulation? Some information-processing systems can internally *simulate* the external behavior of a target object (see BNO, section 2.3). The simulation of a target system consists in representing those of its properties that are accessible to sensory processing, and the way in which they probably develop over time. Some information-processing systems, however, form special cases in that they can also *emulate* the behavior of *another* information-processing system. They do so by internally simulating not only its observable output, but also hidden aspects of its internal information processing itself. Such hidden aspects can consist in abstract properties, like its functional architecture or the software it is currently running. A third possibility, which is of particular interest from a philosophical perspective, is *self-directed* emulation. Self-modeling is that special case, in which **target system and simulating/emulating system are identical**: A self-modeling information-processing system internally and continuously simulates its own observable output as well as it *emulates* abstract properties of its own internal information-processing – and it does so *for* itself.

In short, a self-model is an integrated model *of* the very representational system, which is currently activating it within itself, as a whole. Typically it will possess a bottom-up component driven by sensory input (self-presentation). This input perturbs or modulates the incessant activity of top-down processes continuously generating new hypotheses about the current state of the system (self-simulation), thereby arriving at a functionally more or less adequate internal image of the system's overall, *actual* situation (self-representation). However, the pivotal question is: What justifies treating all these highly diverse kinds of information and phenomenal representational content as belonging to *one* entity?

What bundles these differing forms of phenomenal content is a higher-order phenomenal property: The property of *mineness* (often also called the "**sense of ownership**"). Mineness is a property of particular forms of phenomenal content that, in our own case, is introspectively accessible on the level of inner attention as well as on the level of self-directed cognition. Here are some typical examples of how we, linguistically,

refer to this particular, higher-order phenomenal quality in folk psychological contexts: "I experience *my* leg subjectively as always having belonged to me"; "I always experience *my* thoughts, *my* focal attention and *my* emotions as part of *my own* stream of consciousness" or "voluntary acts are initiated by *myself*".

The phenomenal property of mineness is closely related to the property of **phenomenal *selfhood.*** Again, let us look at some examples of how we frequently attempt to point to the phenomenal content of the internal representational states underlying this property, using linguistic tools from public space: "I am *someone*"; "I experience myself as being *identical* through time"; "the contents of my phenomenal self-consciousness form a coherent *whole"*, "before initiating any intellectual or attentional operations, and independently of them I am already immediately and 'directly' acquainted with the fundamental contents of my self-consciousness."

To sum up, a phenomenal self-model is an integrated representation of the system as a whole. Let us now apply some of our constraints, in order to enrich the concept of a PSM.

3.1. Global Availability of System-related Information

3.1.1. The phenomenology of global availability of system-related information The contents of my phenomenal self-consciousness is directly available, it seems, to a multitude of my mental and physical capacities at the same time. I do experience the general global availability of the contents of my self-consciousness as my own flexibility and autonomy in dealing with these contents, and, in particular, by the subjective sense of immediacy in which they are given to me.

However, it is important to point out three more specific phenomenological characteristics. First, the degree of flexibility and autonomy in dealing with the contents of self-consciousness may vary greatly: emotions, or sensations of pain and hunger, are much harder to influence than for instance the contents of the cognitive self-model. There is a **gradient of functional rigidity**, and the degree of rigidity itself is available for phenomenal experience. Secondly, the phenomenal experience of immediacy is a graded feature as well: Typically thoughts are something that may not even be determined in its full content before spoken out aloud or actually written down on a piece of paper, whereas bodily sensations like pain or thirst are directly given as explicit and "ready-made" elements of the phenomenal self. The human self-model exhibits a continuum between transparency and opacity: The self-constructed character accompanying different contents of the conscious self is highly variable. Thirdly, it is interesting to note that first-order states integrated into the PSM as well as second-order attentional or cognitive states operating *on* these contents are both characterized by the phenomenal quality of "mineness". The conscious contents of your current body image are not experienced as *representational* contents, but are endowed with a phenomenal sense of ownership: at any given time, it is your *own* body. While consciously reasoning about the current state of your body, you will typically be well aware of the representational character of the cognitive constructs emerging in the process while at the same time such thoughts about your current bodily state are characterized by the untranscendable conscious experience of "mineness", by just the same immediate sense of ownership. This is the way in which beings like ourselves experience a representational structure as *integrated* into the PSM.

Conscious human beings do not direct their attention to bodily sensations alone, they can also form thoughts *de se*. The content of *de-se*-thoughts is formed by *my own cognitive states* about *myself*. Reflexive, conceptually mediated self-consciousness makes system-related information cognitively available, and it obviously does so by generating a higher-order form of phenomenal content (Metzinger 2003b). This content, however, does not appear as an isolated entity, but is recursively *embedded* into the same, unified phenomenal whole, into the self-model.

3.1.2. Global availability of self-representational content As noted in section 1.3, the existence of a coherent self-representation for the first time introduces a **self-world-border** into the system's model of reality. System-related information now becomes globally available *as system-related* information, because the organism now has an internal image of itself as a whole, as a distinct entity possessing global features. This in turn forms a necessary pre-condition for the conscious representation of dynamic relations holding between the organism and varying objects in its environment.

3.1.3. Informational/computational availability of system-related information Self-related phenomenal information is equivalent to globally available system-related information. One of the fascinating features of the human self-model is that this information ranges from the molecular to the social. For instance, the self-model is important in processing internal information relevant to elementary bioregulation, i.e., it plays a role in molecular-level self-stabilization (Damasio 1999). It is also important in making information about the fact that the system itself is constantly engaged in information-processing and reality-modeling available to a large number of different **metarepresentational** processes. These higher levels include other-agent modeling.

3.1.4. Global availability of self-related information as a functional property Under a functionalist analysis, a PSM is a discrete, coherent **set of causal relations**. It exerts an important causal influence, not only in differentiating and flexibilizing, but also by **integrating the behavioral profile** of an organism. As one's own bodily movements for the first time become globally available as one's *own* movements, the foundations for agency and autonomy are laid, because the organism now has an internal model of itself *as a whole*. A specific subset of events perceived in the world can now for the first time be treated as systematically correlated *self-generated* events. And the fact that there can be events in the world, which are simultaneously self-generated and self-directed can be discovered and made globally available. The most central aspect of the distinct causal role played by a PSM may consist in later enabling the system to become and treat itself as a second-order intentional system (Dennett, 1981, p. 273-284; Dennett 1987a,b), thereby turning it from a behaving system into an *agent*.

3.2. Situatedness and Virtual Self-presence

Let us now apply the presentationality-constraint to the concept of a globally available self-model. Whatever I experience as the content of my phenomenal self-consciousness, I experience it *now*. In addition, it is not only that a world is present; it is that I am a *present self* within this world. **My own existence possesses temporal immediacy**: a sense of being in touch with myself in an absolutely direct and non-mediated way, which could not be bracketed. If it was possible to subtract the phenomenal content now at issue, I would simply cease to exist on the level of subjective experience.

3.2.1. The phenomenology of self-presence and temporal situatedness
Phenomenal experience does not only consist in "being present". It also consists in "being present *as a self*". Interestingly, there is now a more specific sense of "internality" – i.e., *temporal* internality – that overlaps with the more general sense of internality constituted by the notion of self-representation. Phenomenologically speaking, I am not only *someone*, but also someone who is *situated in a temporal order*. A psychological moment has come into existence, and can now be integrated with autobiographical memory. Human beings can consciously experience the historicity of their own person: The conscious experience of being a self having a past and a future while being currently localized at a specific point within a given temporal order.

3.2.2. *De-nunc*-character of the PSM Even when carrying out a phenomenal self-simulation, e.g., when making plans about my own distant future, or when spontaneously simulating past states of myself, it is always clear that I am making these plans *now* and that I am having these memories *now*. Interestingly, our capacity for mental time travel is never complete. Temporarily, our attention may be fully absorbed by simulational content generating future selves or by recreating the legend of a putative past self, but there is a subtle phenomenal presence of bodily awareness, which is never entirely lost. It anchors us within the phenomenal window of presence generated by the physical system, which we are. In fact, this may be one of the greatest achievements of the human self-model: It *integrates* the representational content constituted by basic, bioregulatory information-processing currently carried out in order to keep the physical condition of the body stable, with higher-order cognitive contents simulating possible states of the organism. It is the self-model, as it were, which bridges the gulf from the actual to the possible, from the bodily to the cognitive. It **links self-representations and self-simulations** by the common phenomenal property of mineness, and the generation of this property depends decisively on the generation of a stable context, provided by temporal internality transparently represented via content *de nunc* (see BNO, p. 555f.).

3.2.3. Self-presence as an informational/computational property At this point I have to return to my old favourite, the virtual reality metaphor. From an epistemological point of view, every self-representation actually is a self-*simulation*. If we look at it from a third-person perspective, then it never truly models or "grasps" the current physical state of the system. This is also true of self-*presentation*: "*Anwesenheit*", the robust phenomenal experience of being present as a self, strictly speaking, is just a form of memory. However, if it approximates the target properties forming the intentional content of its simulation in a functionally adequate manner, if it simulates its own physical dynamics in a good enough way, it may treat such contents as temporally internal. In doing so it can behave as if it was actually fully immersed in the reality that it is simulating. Please note how this is a goal shared with certain technical systems like Virtual Reality-interfaces (see BNO, p. 553f)

3.2.4. Self-presence as a functional property A system continuously modeling itself in a window of presence thereby gains a number of new functional properties. It generates a reference basis for phenomenal self-simulations. For example, autobiographical memories can now be compared and related to the current status of the system. Explicit planning becomes possible. From a teleofunctionalist perspective, self-simulations not co-varying with actual properties of the system can only be turned into

helpful tools (e.g., in forward modeling motor behavior or in making future plans), if a representation of the current state of the system *as* the current state of this system is in existence. Self-modeling within a window of presence achieves precisely this. In BNO (p. 285, 313, 338) I have called this the "Self-Zero-Hypothesis", and there is a corresponding "World-Zero-Hypothesis" concerning the *general* function of consciousness as well (e.g., p. 61).

3.3 Transparency: From System Model to Phenomenal Self

Applying the transparency-constraint to the concept of a conscious self-model is the decisive step in understanding how the conscious experience of *selfhood* can be reductively explained. An active, dynamical "self-model", upon closer inspection, is just a representation of the system as a whole; it is a *system-model*, and certainly not a self. A particularly malicious opponent might even argue that by introducing the concept of a "self-model" I have actually cheated, perhaps accusing me of installing an intuition-pump. It ultimately rests on an equivocation smuggled in by the help of the word "self": A self-*directed* ("reflexive") process of creating an inner picture of the organism as a whole is not the same as a process of internally portraying *a* self. In this sense, a system-model simply is not a *self*-model. Any machine can do self-directed modeling, and, in fact, many machines do this today.

What is needed to determine that a *genuine* experience of *being someone*, of phenomenal selfhood will come into existence? The prereflexive, preattentive experience of *being someone* directly results from the contents of the currently active system-model being *transparent*. Any system acting under a transparent self-model will, if all other necessary conditions for the emergence of phenomenal experience are realized, *by necessity* experience itself as being in direct and immediate contact with itself.

3.3.1. The phenomenology of transparent self-modeling As noted above, we are a systems caught in a *naive-realistic self-misunderstanding*. There are large classes of phenomenal states, in which our self-model is completely transparent, and we do not think or engage in higher-order processes of self-modeling like self-directed attention and cognition. In a certain sense we are "one with ourselves" in such situations. We do not *distance* us from ourselves by generating higher-order self-representational content. Many animals and most human infants may be in this stage. From a phenomenological point of view, it is interesting to now ask what the exact opposite situation would be. The answer is that it never seems to exist: There simply is no conscious self-representation characterized by opaque content in its entirety; cognitive self-reference always takes place against the background of transparent, pre-conceptual self-modeling (Metzinger 2003b). That is to say, there simply are no phenomenal state-classes, in which we experience ourselves as pure, disembodied spirits, not possessing any location in a real temporal order or in physical or behavioral space (see BNO, section 7.2.3). SMT makes a straightforward phenomenological prediction: Were the PSM to become opaque in its entirety, the phenomenal property of selfhood would disappear.

At this point we must briefly return to the threat of circularity mentioned above, to the issue whether transparency *really* is a necessary condition for phenomenality. The terminological conventions we will have to adopt are in part interest-relative, because they depend on our epistemic goals: If we want to understand how ordinary conscious

experience becomes a *subjective* phenomenon by being tied to a first-person perspective originating from a robust phenomenal self, then we have to assume a stable transparent partition in the self-model. We need the transparency constraint, because only the transparency of the PSM gives us a robust phenomenal self. If our explanatory scope is wider, by including selfless phenomenal states as they can be found in certain spiritual experiences or in severe psychiatric conditions like depersonalization, then we may not need the transparency constraint. For instance, there could be **phenomenologically non-subjective state-classes** in which the system only operates under an integrated, but phenomenally opaque *system*-model. Imagine a situation in which the lucid dreamer would also recognize *herself* as being a dream character, a simulated self, a representational fiction – a situation, in which the dreaming system, as it were, became lucid *to itself*. From a deeper philosophical perspective, such states might be highly relevant. In BNO (p. 566) I have introduced the term "system-consciousness" to capture this possibility. The conceptual decision to be made is whether we want to call such conscious states "experiences". If we do, transparency is neither a necessary condition for subjectivity, nor for phenomenality. For analytical philosophers like me one central problem will always be that autophenomenological reports about such states contain an inherent logical fallacy (a "performative self-contradiction"): How can you coherently report about a selfless state of consciousness by referring to your *own*, autobiographical memory?

Again, in order to do justice to the real phenomenology, one has to admit that the property of phenomenal transparency at issue is not an all-or-nothing phenomenon, but may be distributed across varying parts of the human self-model to differing degrees. In general, the bodily self-model is fully transparent, while high-level cognitive processes like reasoning are phenomenally opaque. However, one of the particularly interesting features about the phenomenology of human self-consciousness is that there are aspects, e.g., certain emotional processes, which on the level of subjective experience may *oscillate* back and forth between transparency and opacity multiple times. This is particularly evident in social relationships. Subjective experiences of trust, jealousy or mild paranoia are interesting examples. The phenomenology of transparent experience is the phenomenology of not only knowing, but of also knowing *that* you *know* while you know; opaque experience is the experience of knowing while also (non-conceptually, attentionally) knowing that you may be *wrong*. In trusting another human being, a certain part of your emotional self-model has a direct and perception-like quality: You just *know* that you know that a certain fellow human being is trustworthy and this conscious experience is accompanied by a maximal sense of certainty. If that person disappoints you, not only your phenomenal model of that person suddenly changes, but at the same time a certain internal de-coherence or disassociation in your own self-model is created: You realize that your emotional state of trust was just a *representation* of social reality, and in this case it was a misrepresentation. It becomes opaque. Phenomenologically, a vehicle/content distinction is introduced where there previously was none. Again, we find that the emotional level of self-representation takes a middle position between the extremes.

The transparent partition of the conscious self-model is of particular importance in generating the phenomenal property of selfhood, in making processes of sensorimotor integration globally available, and in generating an internal user surface for motor

control. However, if it had not been for the opaque partition of my self-model, I could not have written this Précis. What makes the conscious self-model of human beings so unique, and so enormously successful as a representational link between biological and cultural evolution, is the fact that it violates the principle of autoepistemic closure. The fact of us possessing an opaque part of our self-model allows us to *conceive of* the possibility of an appearance/reality distinction not only for our own perceptual states, but for the contents of self-consciousness as well. It allows us to distance us from ourselves by critically assessing the content of any PSM and—by opaque simulation—it allows us to conceive of certain possibilities, for instance the epistemological possibility that every phenomenal representation might actually be a *simulation*, if viewed from an objective third-person perspective (see BNO, Chapter 2). It also allows us to, for the first time, conceive of the possibility that every phenomenal *self*-representation might actually be a self-*simulation*. Such cognitive discoveries, of course, do not yet change the fundamental architecture of our phenomenal space.

3.3.2. Transparency as a property of self-representation The representational vehicle of your conscious self-experience is a certain process in your brain, a complex neural activation pattern. This process of self-representation is not consciously experienced by you. It is not globally available for attention and it is transparent in the sense of you currently looking through it. In this special case, what you are looking at is *yourself*: What you are seeing and feeling *onto* is its self-representational content, e.g. the existence of your hands, here and now, given through a multitude of internal as well as external sensory channels. This content is an abstract property of the concrete self-representational state, of the currently active self-representatum in your head.

Please recall that there are at least two kinds of mental content: The *intentional* content and the locally supervening *phenomenal* content of the self-representation (See Section 2.3. above). A bodiless brain in a vat could certainly enjoy the *phenomenal* experience of holding a paper like this one in its *own* hands right now. The phenomenal content of your bodily self-representation is entirely determined by internal properties of your brain. If, while reading these sentences, you actually *are* a brain in a vat, then you are, as it were, not any more "intentionally" or "epistemically" looking through a state in your head onto your hands, but only onto this state itself – without *this* fact being globally available to you on the level of phenomenal representation.

The phenomenal property of selfhood is constituted by transparent, non-epistemic self-representation, and it is on this level of representationalist analysis that the refutation of the corresponding **phenomenological fallacy** becomes truly radical, because it has a straightforward ontological interpretation: no such things as selves exist in the world.

3.3.3. Transparent self-modeling as an informational/computational property Viewed as a computational strategy, transparent self-modeling drastically reduces computational load. In particular, it prevents the system from being caught up in an infinite regress of self-modeling. It is important to note that self-modeling, in terms of its logical structure, is an infinite process: A system that would model itself as currently modeling itself would thereby start generating a chain of nested system-related mental content, an endless progress of "self-containing", of conscious self-modeling, which would quickly devour all its computational resources and paralyze it for all practical purposes. It thus has to find an efficient way to break the reflexive loop. One simple and

efficient way to interrupt a circular structure is by introducing an untranscendable *object*: My hypothesis is that the phenomenon of *transparent* self-modeling developed as an evolutionary viable strategy, because it constituted a reliable way of making system-related information available without entangling the system in endless internal loops of higher-order self-modeling. I call this the **"principle of necessary self-reification"**: What we experience as our phenomenal self on the most fundamental level is precisely the transparent representational object that blocks self-representational loop. Interestingly, this process, on closer inspection, has not resulted in object-formation, but in *subject*-formation (see BNO, section 6.5).

3.3.4. Transparent self-modeling as a functional property Systems operating under a transparent self-model constantly assume their own existence as an individual, coherent entity. They become *realists—naive realists—*about themselves, and this will obviously have functional consequences. I like to put this point by saying that the possession of a transparent PSM makes a system maximally *egotistic*.

It is, in terms of new functional properties brought about by a transparent self-model, highly interesting to note that, a) the self-model is an entity entirely located on the sub-personal level of description while at the same time b), it is the decisive link in enabling **personal-level communication** between human beings and within larger groups. You *become* a person by having the right kind of subpersonal self-model, one that functionally allows you to enter mutual relationships of acknowledging each other's personhood in a social context.

3.3.5. Neural correlates of transparent self-modeling Again, not much is presently known about the neural underpinnings of the transparent self-model in humans. However, let me point out that the concept of a transparent PSM bears at least some similarity to Antonio Damasio's notion of the "core-self" (see Damasio 1999, 2000).

3.4. Convolved Holism and the Phenomenal Self

The sub-region of reality which is phenomenally experienced *as internal*, i.e. the self-model, possesses a holistic character, and this holism is pervasive, because it also holds for many different forms of ever changing phenomenal content out of which it is composed.

3.4.1. Convolved holism as a phenomenological feature of self-consciousness The phenomenal self constitutes a subglobal whole (a "world within the world"). The concrete wholeness of my own self is characterized by a multitude of internal part-whole relationships. However, any realistic phenomenology will have to do justice to the fact that this hierarchy is a highly flexible, "liquid" hierarchy. As the focus of my introspective attention wanders, the overarching wholeness is never threatened. What continuously changes, however, is the way in which bodily, emotional and cognitive contents of experience are integrated, and transiently nested into each other.

3.4.2. Convolved holism as an informational/computational strategy for self-representation System-related information represented in a holistic format is coherent information which is available to the system as a *single* possible object of intended cognition and focal attention. At the same time, information integrated into a convolved holistic self-model generates an *internal* kind of interdependence: As single features, e.g.,

background emotions and cognitive states, directly influence each other within this model, it provides a new way of representing the complex causal structure governing the internal dynamics of the system itself.

3.5. Dynamics of the Phenomenal Self

3.5.1. Phenomenology of the dynamic self Whatever my true nature is, I am an entity which undergoes changes. I can introspectively discover all of the phenomenological aspects of time-experience (BNO, section 3.2.5) in myself: There is the simultaneity of bodily sensations; there is succession and seriality as paradigmatically experienced in conscious reasoning; I experience myself as a part of and directly in touch with reality by being a *present* self; and, finally, the phenomenology of self-consciousness is certainly characterized by a strong element of duration. This last aspect is particularly interesting, because the coherence and duration of the phenomenal self is, as a matter of trivial fact, highly discontinuous, for instance in being repeatedly and reliably interrupted by phases of deep and dream sleep. It is the invariance of bodily self-awareness and autobiographical memory, which constitute the conscious experience of an enduring self. The conceptual reification of what actually is a very instable and episodic process is then reiterated by the phenomenological fallacy[3] pervading almost all folk-psychological, and a large portion of philosophical discourse on self-consciousness. But it is even phenomenologically false: We are not things, but processes.

3.5.2. Dynamicity as a property of phenomenal self-representation One of the important ideas inherent in dynamicist cognitive science is not to conceive of intentionality as a rigid, abstract relation pointing from a subject to an intentional object, but as a dynamical physical process. In the same way, *reflexivity* will now not be a rigid, abstract relation in which a subject stands to itself, but a constructive, dynamical physical process generating a constantly updated self-model.

3. 6. Perspectivalness

There is a way of creating something resembling perspectivalness even *within* the self-model, namely in higher-order variants of cognitive self-consciousness. For the purposes of this Précis, this is not an issue of central interest. What is more important is how a transparent self-model can function as the *origin* of the consciously experienced first-person perspective, by becoming the more invariant part of a yet more complex form of phenomenal content. I will briefly sketch it below, in the last section.

4. The PMIR: The Consciously Experienced First-Person Perspective

Any computational system operating under a world-model centred by coherent self-model has introduced the most fundamental partitioning of its informational space possible: the differentiation between the processing of environment-related and system-related information.

A phenomenal subject, as opposed to a mere phenomenal self, is a model of the system as *acting and experiencing*. What is needed is a theory about how the intentionality-relation, the relation between subject and object is *itself* depicted on the level of conscious experience. What is needed is a theory about what in previous publications I have introduced as the "phenomenal model of the intentionality-relation" (Metzinger 1993, p. 128 pp.; 2000b, p. 300, 2005).

4.1. The Concept of a PMIR: A Short Representationalist Analysis

The phenomenal model of the intentionality relation (PMIR) is a conscious mental model[4], and its content is an ongoing, episodic subject–object-relation. Phenomenologically, a PMIR typically creates the experience of a self in the act of knowing, of a self in the act of perceiving - or of a *willing* self in the act of intending and acting.
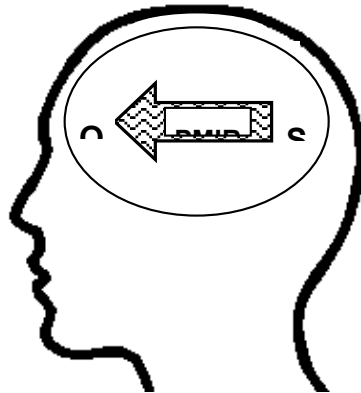

Figure 1 The Phenomenal Model of the Intentionality-Relation (PMIR): A subject-component (**S**; the PSM, an internal, conscious model of the system as a whole,) is phenomenally represented as directed at an object-component (**O**; the "intentional object"). In conscious volition **O** always is a *goal-component*, for example, an allocentric representation of a successfully terminated bodily action.

The notion of a PMIR must be clearly separated from the *classical* concept of intentionality, as it can be found in Franz Brentano (1874). Good Old Fashioned Intentionality (GOFI) is a relation between a mental act and an object-component, which is mentally contained in the mode of "intentional inexistence".
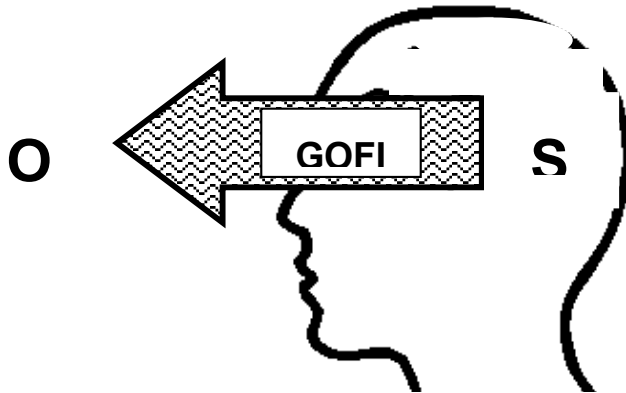
Figure 2 Good Old Fashioned Intentionality (GOFI): A subject-component (**S**; the "mental act") is directed at an object-component (**O**; the "intentional object"). As **O** does not necessarily exist, **GOFI** is a non-physical relation.

What I want to draw attention to is a point, which has been frequently overlooked in the past: The classical intentionality-relation can *itself* form the content of a conscious mental representation. In beings like us, there exists a phenomenal model *of* the intentionality relation. 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 non-human animals are intentional systems, but that their nervous systems do not allow them to ever become aware of this fact. In any case, it is important to note how, in our own case, GOFI can itself be a form of phenomenal content: My central point is that **we do not only represent individual objects, 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 consciousness is experienced as involving a first-person perspective.

4.2. What is the Function of a PMIR?

Phenomenal mental models 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. A phenomenal **model of transient subject–object relations** makes an enormous amount of new information available for the system: All information related to the fact that it is currently perturbed by perceptual objects, that certain cognitive states are currently occurring in itself, e.g., to the fact that certain abstract goal representations are currently active, that there are a number of concrete self-simulations connecting the current system-state with the state the system would have if this goal state would be realized; allowing for selective behavior and the information that it is a system capable of manipulating its own sensory input, e.g., by turning its head and directing its gaze to a specific visual object. 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.

A PMIR also allows for a dynamical representation of transient subject-object relations, and thereby makes a **new class of facts** globally available. If the capacity to iterate the construction of a PMIR is given (i.e., to point a second-order arrow at a first-order arrow, to turn one PMIR into the object-component of another one), two entirely new forms of intelligence emerge, because the system can now flexibly and selectively react to two entirely new classes of facts:

Introspective Intelligence

Agency: The system can become aware of the fact that it *has* a will and that it *is* an agent (= selectively directed at goal-states).

Attentional subjectivity: The system can become aware of the fact that it *has* selective, high-level attention. This enables more complex forms of learning and epistemic autoregulation.

Reflexive self-consciousness: If it has the capacity for conceptual thought, the system can mentally represent the fact *that* it has a first-person perspective. This makes the transition from phenomenal to cognitive subjectivity possible.
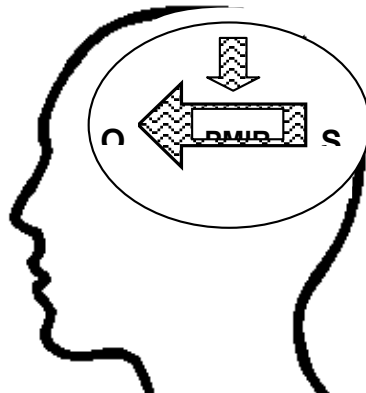

Figure 3: *Introspective Intelligence*: A second-order **PMIR** is directed at a first-order **PMIR** as its object-component. The underlying principle of relational coding is iterated. In the special case of conscious second-order volition, the first-order **PMIR** is an ongoing representation of the system as currently directed at a goal-component. The volitional second-order **PMIR**, within certain temporal boundary conditions, can allow you to *terminate* the initiation of an action, or to consciously represent the fact that you *want to want* something.

Social Intelligence

Other agent-modeling: The system can become aware of the fact that other systems have a first-person perspective too. This permits action coordination and cooperative behavior.

Mind-reading: The system can internally simulate external PMIRs. It can develop an empathic understanding of other agents.

High-level intersubjectivity: Systems with *cognitive* PMIRs can mutually acknowledge each other as persons. This enables the emergence of *normative* intersubjectivity and complex societies.
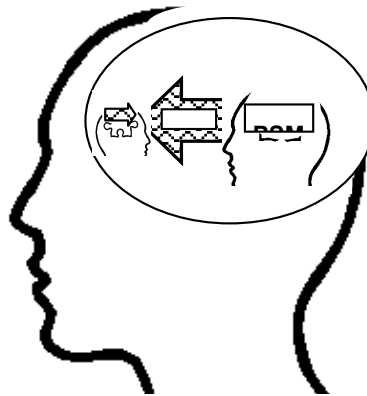

Figure 4 Social Intelligence: A second-order **PMIR** is directed at a first-order **PMIR**, which as been *integrated into the model of another agent*. The underlying principle of relational coding is now iterated in the social domain. In the special case of consciously experienced sociovolitional cognition, the second-order **PMIR** is an ongoing representation of the system as currently directed at *the intention of another agent*. The brain integrates a volitional first-order **PMIR** into a model of another agent, either as currently perceived in the environment or as mentally simulated. Again, the functional significance of this architecture is that a *new class of facts* can be consciously integrated, facts having to do with the actual or potential existence of other goal-directed beings, other intentional agents in the environment.

And this may be a more general insight pertaining to *all* the different forms of phenomenal content discussed in BNO or in this Précis: Phenomenal models are neurocomputational instruments, which endow an organism with new functional properties, because they make certain types of facts globally available for processing within a virtual window of presence, by simultaneously generating a single, integrated context and a temporal frame of reference. The PSM of *Homo sapiens* was not only

special in that, by becoming embedded into the global model of reality, it allowed us to become aware of our own existence. Through its opaque partition it also permitted us to become aware of our own existence *as representational systems operating under an individual first-person perspective*. We could now conceptually self-ascribe this property to us, linguistically communicate it, and thereby opened the door which makes the transition from biological into cultural evolution possible.

Notes

1. The representandum is the *object* of representation. The representatum is the concrete internal *state* carrying information related to this object, at a given point in time. Representation is the *process* by which the system as a whole generates this state. Therefore, the "representatum" is a time-slice of the ongoing, physically realized process of representation.

2. Please note that, in the interest of brevity, I am here simplifying matters perhaps too much. In BNO (p. 36), 4 different concepts of "introspection" are distinguished.

3. This fallacy consists in the unjustified use of an existential quantifier within a psychological operator: If I look into a red flash, close my eyes and then experience a green afterimage, this does not mean that now a non-physical object possessing the property of "greenness" has emerged. Cf. an early formulation by Place, 1956, section V: "This logical mistake, which I shall refer to as the 'phenomenological fallacy' is the mistake of supposing that when the subject describes his experience, when he describes how things look sound, smell, taste or feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen, usually referred to in the modern psychological literature as the 'phenomenal field'."

4. The concept of a "phenomenal mental model" (Metzinger 2003a, chapter 3) is loosely connected to a theory of mental representation developed by Philip Johnson-Laird (see, e.g., 1983, 2001), which it extends to the phenomenological domain by imposing and additional set of constraints. Models are representational entities, the fact that they are models *of* something does not emerge on the phenomenal level. Unconscious self-models ("SMs") and non-phenomenal precursors of a representation of the intentionality-relation ("MIRs") could exist. However, *phenomenal* models possess a number of special functional/neurocomputational features which are relevant for understand the volitional process (and its disorders) as a whole.

References

Andrade, J. (2000). Using anesthetics to assess the role of conscious processes in learning. In Metzinger 2000a.

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.

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

Chalmers, D.J. (1997). Availability: The cognitive basis of experience? In Block, Flanagan and Güzeldere 1997.

Chalmers, D.J. (2000). What is a neural correlate of consciousness? In Metzinger 2000a.

Conant, R.C., and Ashby, W.R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science* 2: 89-97. Reprinted in G. J. Klir, ed., (1991), *Facets of System Science*. New York: Plenum Press.

Cruse, H. (1999). Feeling our body - The basis of cognition? *Evolution and Cognition* 5: 162-73.

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

Damasio, A.R. (2000). A neurobiology for consciousness. In Metzinger 2000a.

Dennett, D.C. (1981[1978]). *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MA: MIT Press.

Dennett, D. C. (1987a). *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, D. C. (1987b). Intentional systems in cognitive ethology: The Panglossian Paradigm defended. In Dennett 1987a. First published in *Brain and Behavioral Sciences* 6: 343–390.

Edelman, G.M., and Tononi, G. (2000a). Reentry and the dynamic core: neural correlates of conscious experience. In Metzinger 2000a.

Edelman, G.M. & Tononi, G. (2000b). *A Universe of Consciousness. How Matter becomes Imagination*. New York: Basic Books.

Engel, A.K., and Singer, W. (2000). Binding and the neural correlates of consciousness. *Trends in Cognitive Sciences* 5: 16-25.

Flohr, H. (2000). NMDA-receptor-complex-mediated computational processes as a candidate for the NCC. In Metzinger 2000a.

Fodor, J.A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Frank, M. (1991). *Selbstbewußtsein und Selbsterkenntnis*. Stuttgart: Reclam.

Franks, N.P., and Lieb, W.R. (2000). An assessment of the role of NMDA receptor function in consciousness: What can we learn from the mechanisms of general anesthesia? In Metzinger 2000a.

Hardcastle, V.G. (2000). How to understand the N in NCC. In Metzinger 2000a.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (2001) Mental models and deduction. *Trends in Cognitive Sciences*, 5, 434-442.

LaBerge, S., and Gackenbach, J. (2000). Lucid dreaming. In Cardeña, Lynne and Krippner 2000.

Leopold, D.A., and Logothetis, N.K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences* 3: 254-64.

Llinás, R., and Ribary, U. (1998). Temporal conjunction in thalamocortical transactions. In H.H Jasper, L. Descarries, V.F Castellucci, and S. Rossignol, eds., *Consciousness: At the Frontiers of Neuroscience*. Advances in Neurology, 77. Philadelphia, PA: Lippincott-Raven.

Llinás, R., Ribary, U., Contreras, D., and Pedroarena, C. (1998). The neuronal basis for consciousness. *Philosophical Transactions of the Royal Society of London* B 353: 1841-9.

Llinás, R., Ribary, U., Joliot, M., and Wang, X.-J. (1994). Content and context in temporal thalamocortical binding. In Buzaki *et al*. eds., *Temporal Coding in the Brain*. Berlin: Springer.

Llinás, R.R., and Paré, D. (1991). Of dreaming and wakefulness. *Neuroscience* 44: 521-35.

Llinás, R.R., and Ribary, U. (1992). Rostrocaudal scan in human brain: A global characteristic of the 40-Hz response during input. In Basar and Bullock 1992.

Llinás, R.R., and Ribary, U. (1993). Coherent 40 Hz oscillation characterizes dream state in humans. *Proceedings of the National Academy of Science USA* 90: 2078-81.

Llinás, R.R., and Ribary, U. (1994). Perception as an oneiric-like state modulated by the senses. In Koch and Davies 1994.

McGinn, C. (1989). *Mental Content*. Oxford: Oxford University Press.

McGinn, C. (1991). *The Problem of Consciousness: Essays toward a Resolution*. Oxford: Basil Blackwell.

Metzinger, T. (1993). *Subjekt und Selbstmodell*. Paderborn: mentis.

Metzinger, T. (1995a). *Conscious Experience*. Thorverton: Imprint Academic & Paderborn: mentis.

Metzinger, T. (1995b). Faster than thought. holism, homogeneity and temporal coding. In Metzinger 1995a.

Metzinger, T. (2000a). *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.

Metzinger, T. (2000b). The *subjectivity* of subjective experience: A representationalist analysis of the first-person perspective. In Metzinger 2000a. Revised version (2004) in *Networks*, 3-4: 33-64.

Metzinger, T. (2004a). *Being No One. The Self-Model Theory of Subjectivitity.* Cambridge, MA: MIT Press.

Metzinger, T. (2003b). Phenomenal transparency and cognitive self-reference. *Phenomenology and the Cognitive Sciences*, **2**, 353-393.

Metzinger, T. (2005, in press). Conscious volition and mental representation: Towards a more fine-grained analysis. In N. Sebanz und W. Prinz, eds., Disorders of Volition. Cambridge, MA: MIT Press.

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, 549-71.

Parvizi, J., and Damasio, A. (2001). Consciousness and the brainstem. *Cognition* 79: 135-159.

Place, U.T. (1956). Is consciousness a brain process? *British Journal of Psychology* 47: 44-50.

Pöppel, E. (1988). *Mindworks: Time and Conscious Experience*. New York: Hartcourt Brace Jovanovich.

Pöppel, E. (1994). Temporal mechanisms in perception. *International Review of Neurobiology*, **37**, 185-202.

Pöppel, E. (1995). Homogeneity of space and continuity of time: Necessary prerequisites of perception? In H.J. Herrmann, D.E. Wolf & E. Pöppel (eds.), *Supercomputing in Brain Research: From Tomography to Neural Networks*. Singapore: World Scientific Publishing.

Ruhnau, E. (1995). Time-Gestalt and the observer. In Metzinger 1995a.

Singer, W. (2000). Response synchronization: A universal coding strategy for the definition of relations. In M.S. Gazzaniga (ed.), *The New Cognitive Neurosciences. 2 ed.* Cambridge MA: MIT Press.

Singer, W. (2004). Synchrony, oscillations, and relational codes. In L.M. Chalupa & J.S. Werner, eds., *The Visual Neurosciences*. Cambridge, Massachusetts: The MIT Press.

Singer, W. (2005). Large scale temporal coordination of cortical activity as prerequisite for conscious experience**.** In M. Velmans, ed., *Blackwell Companion to Consciousness*. Oxford: Basil Blackwell.

Salthe, S.N. (1985). *Evolving Hierarchical Systems*. New York: Columbia University Press.

Tononi, G., and Edelman, G.M. (1998a). Consciousness and complexity. *Science* 282: 1846-51.

Tononi, G., and Edelman, G.M. (1998b). Consciousness and the integration of information in the brain. In H.H. Jasper, L. Descarries, V.F. Castellucci and S. Rossignol, eds., *Consciousness: At the Frontiers of Neuroscience*. Advances in Neurology, **77.** Philadelphia, PA: Lippincott-Raven.

Tononi, G. (2003). Consciousness differentiated and integrated. In A. Cleeremans, ed., *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford: Oxford University Press.

Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb : Phase synchronization and large-scale integration. *Nature Neuroscience* 2: 229-239.

von der Malsburg, C. (1997). The coherence definition of consciousness. In M. Ito, Y. Miyashita and E. T. Rolls, eds., *Cognition, Computation, and Consciousness*. Oxford: Oxford University Press.

Yates, J. (1985). The content of awareness is a model of the world. *Psychological Review* 92: 249-84.

Zeki, S., and Bartels, A. (1998). The autonomy of the visual systems and the modularity of conscious vision. *Philosophical Transactions of the Royal Society of London* B 353: 1911-4.