 Mind & Matter Vol. 3(1), pp. 57–84

Out-of-Body Experiences as the Origin of the Concept of a "Soul"

Thomas Metzinger
Frankfurt Institute for Advanced Studies
Frankfurt/Main, Germany

and

Department of Philosophy
Johannes Gutenberg-University at Mainz, Germany

Abstract

Contemporary philosophical and scientific discussions of mind developed from a "proto-concept of mind", a mythical, traditionalistic, animistic and quasi-sensory theory about what it means to have a mind. It can be found in many different cultures and has a semantic core corresponding to the folk-phenomenological notion of a "soul". It will be argued that this notion originates in accurate and truthful first-person reports about the experiential content of a specific neurophenomenological state-class called "out-of-body experiences". They can be undergone by every human being and seem to possess a culturally invariant cluster of functional and phenomenal core properties similar to the proto-concept of mind. The common causal factor in the emergence and development of the notion of the soul and the proto-concept of mind may consist in a yet to be determined set of properties realized by the human brain, underlying the cluster of phenomenal properties described in the relevant first-person reports. This hypothesis suggests that such a neurofunctional substrate led human beings at different times, and in widely varying cultural contexts, to postulate the existence of a soul and to begin developing a theory of mind.

1. The Proto-Concept of Mind

What is the "proto-concept of mind"? In many cultures we simultaneously find prescientific theories about a "breath of life", e.g., the Hebrew *ruach*, the Arabic *ruh*, the Latin *spiritus*, the Greek *pneuma* or the Indian *prana* and the five *koshas*, respectively, etc. (for historical details and further references see Verbeke 1974, Schrott 1974). Typically this is a spatially extended entity, keeping the body alive and leaving it during phases of unconsciousness and after death. It has a material aspect, though more subtle than that of the physical body. We are confronted with an almost ubiquitous idea of what mind actually is, which in all its

many variations still is a *sensory-concrete* idea of the mental as something that integrates parts, mostly of physical organisms, but sometimes, in a wider sense, also of societies and groups of human beings.

In occidental philosophy of mind this proto-concept of mind has developed through innumerable stages, starting from the pneumatology of Anaximenes in the 6th century B.C., through Diogenes of Apollonia and the Aristotelian distinction between breathed air and psychic pneuma (which may perhaps count as the first attempt at a naturalist theory of mind in Western philosophy). This development then continued through alchemist theories of controlling nature by controlling mind and the Neoplatonists, for whom the pneuma was an aureole covering the soul and protecting it from contact and contamination by material objects, on towards Christian philosophy, which finally *denaturalized* and *personalized* the concept of mind (for details and further references see Oeing-Hanoff *et al.* 1974). In this way the Western history of the concept of mind can be read as a history of a continuous differentiation of a traditionalistic, mythical, sensory proto-theory of mind which gradually led to a more and more *abstract* principle. Finally, culminating in Hegel, mind is conceived as devoid of all spatial and temporal properties.

2. The Folk-Phenomenological Concept of a Soul

What is folk-phenomenology? Just like *folk-psychology*, generally it is a naive, prescientific way of speaking about the contents of our own minds – *folk-phenomenology* is a way of referring specifically to the contents of conscious experience, as experienced from the first-person perspective. It generates no or little theoretical progress (just as Churchland (1981) argued for folk-psychology), and is characterized by an almost all-pervading naive realism.

However, in everyday life, folk-phenomenology works remarkably well – at least it seems to. All of us are experienced folk-phenomenologists, because all of us are used to self-ascribe certain phenomenal properties when reporting the content of our phenomenal states to our fellow human beings. In non-scientific contexts, we all know what we mean by a "soul": Our soul is the innermost and essential part of ourselves; it is the prime candidate for the "true self"; it is the phenomenal locus of identity; it bears a deep relation to the emotional layers of our self-model, to the emotional core of our personality. For many of us it is something of which we secretly hope that it may survive physical death, because it is not identical to our body. Folk-phenomenology follows Cartesian intuitions, and the deeper reason for this fact may be that its ontology is mirrored in the representational architecture of the human self-model (Metzinger 2003a, section 6.4.1).

3. Out-of-Body Experiences: Phenomenology, Psychology, and Neuroscience

Could there be an integrated kind of bodily self-consciousness, maybe a mobile body fully available for volitional control, or a paralyzed body which in its entirety is a phenomenal confabulation – in short, a *hallucinated* and *bodily* self at the same time? Is it conceivable that something like a "globalized phantom-limb experience", the experience of a phantom body could emerge in a human subject? The answer is yes. There is a well-known class of phenomenal states in which the experiencing person undergoes the untranscendable and highly realistic conscious experience of leaving his or her physical body, usually in the form of an ethereal double, and moving outside of it.

These states correspond to a class (or at least a strong cluster) of intimately related phenomenal models of reality characterized by a *visual representation* of one's own body from a perceptually impossible, externalized third-person perspective (e.g., lying on a bed or the road below oneself) plus a *second representation* of one's own body, typically (but not in all cases) freely hovering above or floating in space. This second body-model is the locus of the phenomenal self. It forms the "true" focus of one's phenomenal experience and also functions as an integrated representation of all kinesthetic qualia and all non-visual forms of proprioception. Such experiences are called out-of-body-experiences or OBEs. Let us take a closer look at this highly interesting class of phenomenal states.

3.1 Phenomenology and Representational Content

OBEs frequently occur spontaneously while falling asleep, following severe accidents, or during surgical operations. At present it is not clear whether the concept of an OBE possesses one clearly delineated set of necessary and sufficient conditions. The concept of an OBE may in the future turn out to be a cluster concept constituted by a whole range of diverging (possibly overlapping) subsets of phenomenological constraints, each forming a set of sufficient, but not necessary, conditions. On the other hand the OBE clearly is something like a phenomenological *prototype*. There is a core to the phenomenon, as can be seen from the simple fact that many readers will have already heard about in one way or another.

One can offer a representationalist analysis of OBEs by introducing the concept of a "phenomenal self-model" (PSM; for more on the concept of a PSM, see Metzinger 2003a). A PSM is an integrated, conscious representation of the organism *as a whole*, including not only its spatial features, but also those of its own psychological properties to which it has access. An important feature of the human PSM is that it is almost

entirely *transparent*. This means that we, as the organisms activating the PSM in their own central nervous system, cannot recognize it *as* a model: We become naive realists with regard to its content, the transparent representational content of the PSM is simply what we experience and later refer to as "our" conscious self.

Given this conceptual background, we can analyze OBEs a class of deviant self-models. On the level of conscious self-representation a prototypical feature of this class of deviant phenomenal self-models seems to be the coexistence of (a) a more or less veridical representation of the bodily self, from an external visual perspective, which does *not* function as the center of the global model of reality, and (b) a second self-model, which according to subjective experience largely integrates proprioceptive perceptions – although, interestingly, weight sensations only to a lesser degree – and which possesses special properties of shape and form that may or may not be veridical. Both models of the experiencing system are located within the same spatial frame of reference (this is why they are *out-of*-body-experiences). This frame of reference is *egocentric*.

The first interesting point seems to be that the second self-model always forms the subject-component of what I have elsewhere called the "phenomenal model of the intentionality-relation" (PMIR[1]; see Metzinger 2003a, section 6.5, Metzinger 2005). The PMIR itself – the first-person perspective as consciously experienced, the ongoing relationship between subject and object as *itself* phenomenally represented – is almost invariably portrayed as of a perceptual, i.e., visual, nature. Phenomenologically, you simply *see* yourself. If, for instance, after a severe accident, you find yourself floating above the scene viewing your injured body lying on the road beside your car, there is a *perceived* self (the "object-component", which, technically speaking, is only a *system*-model, but not a *subject*-

[1]The concept of a PMIR refers to the phenomenological observation that human beings do not only represent, but that they also *co-represent the representational relation itself* – and often consciously experience this very fact while doing so. Therefore, the PMIR is here conceived of as a conscious mental model, and its content is an ongoing, episodic subject–object-relation. 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 color 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. Such relationships can be perceptual, attentional, cognitive, or volitional. The content of consciousness never is a mere object, it always is a *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.
The notion of a PMIR as used here should not be confused wit Stanford's (1976) concept of a "psi-mediated instrumental response".

model), invariably formed by a more or less accurate visual representation of your body from an exteriorized perspective, and a *perceiving* self (the "subject-component", the phenomenal self-model or PSM, i.e., the current *self-* or *subject*-model), hovering above the scene.

Both self-models are integrated into one overall global model of reality, which is centered on the second self-model. The second self-model can either be one of a full blown agent, i.e., endowed with the characteristic form of phenomenal content generating the subjective experience of agency (see Metzinger 2003a, section 6.4.5), or only what Irwin (1985, p. 310) has aptly called a "passive, generalized somaesthetic image of a static floating self". However, before entering into a brief representationalist analysis of OBEs, let us first take a quick detour and look at some more frequent, real-world phenomenological cases.

Have you ever had the following experience? The bus to the train station had already been late. And now you have even queued up in a line at the wrong ticket counter! Nevertheless you manage to reach your train just in time, finding an empty compartment and, completely exhausted, drop into the seat. In a slightly unfocused and detached state of mind you are now observing the passengers sitting in the train on the other side of the platform. Suddenly you feel how your *own* train starts to move, very slowly at first, but accompanied by a continuous acceleration, which you can feel in your own body. Two or three seconds later, with the same degree of suddenness, your bodily sensation disappears and you become aware that it actually is the *other* train, which has now started to slowly leave the train station (see also Metzinger 1993, p. 185f).

Such an experience is a very rudimentary form of an OBE, a hallucinated bodily self. The center of your global model of reality was briefly filled by a kinesthetic and proprioceptive hallucination, a non-veridical model of the weight and acceleration of your body, erroneously activated by your brain. The dominating visual model of your environment, largely formed by the input offered through the "picture frame" of the train window, was underdetermined. In the special input configuration driving your visual system it allowed for two coherent interpretations: either it is the other train, or it is the train in which you are sitting, which has just started to move. The visual model of reality allowed for two equally consistent interpretations. At the same time there was a state of general physical and emotional arousal, accompanied by an unconscious state of expectancy about what is very likely going to happen next, and very soon.

The information-processing system, which you *are*, selected one of the two possible interpretations in accordance with constraints imposed by a preexisting internal context and, as a system that always tries to maximize overall coherence, "decided" to simultaneously activate a suitable self-model, one that can be integrated into the new phenomenal model of the world without causing any major problems. Unfortunately, the

chosen model of the world was wrong. Therefore, the activation of the accompanying kinesthetic-proprioceptive self-model led the system into a very brief hallucinatory episode. Since transparent models of reality and the self are always fully interpreted and intranscendable for the system currently operating *under* them, a hallucinated bodily self ensued. Its content was the content of a phenomenal self-simulation, activated by an erroneous automatism leading the system astray, while not being recognized as such. A possibility was depicted as a reality. As the dominant visual model of reality is being updated, this briefly "deviating" form of self-modeling leading to the subjective experience of a real body being slowly accelerated is immediately terminated – and with a mild degree of irritation or amusement you recognize that you have just fooled *yourself*.

This may count as the minimal case of a phenomenal self-simulation fulfilling no proper function for the system – in this case leading to a partially empty, illusory experience of the body as a whole and in motion. It does not satisfy the adaptivity-constraint (it has no function for the system as a whole; see Metzinger 2003a, section 3.2.11), and its most striking neurophenomenological feature is the internal emulation of kinesthetic "motion" qualia, of a form of sensory content we normally take to be strictly stimulus-correlated. (See Figs. 1 and 2 for illustrations.) The solution to this problem is to acknowledge that visual kinesthetic information, generally richer than mechanical kinesthetic information, can overrule the second type in cases of conflict, because vision "is not only an exteroceptive sense, as is classically assumed, it is also an autonomous kinesthetic sense." (Lishman and Lee 1973, p. 294). What is still missing in this introductory case study is a stable, exteriorized visual perspective of the physical body. Let us now proceed to look at two classical phenomenological descriptions of spontaneous OBEs in an ordinary non-pathological context (Waelti 1983, p. 25; English translation TM):

> I awoke at night – it must have been at about 3 a.m. – and realized that I was completely unable to move. I was absolutely certain I was not dreaming, as I was enjoying full consciousness. Filled with fear about my current condition I had only one desire, namely to be able to move my body again. I concentrated all my will-power and tried to roll over to one side: Something rolled, but not my body – something that was me, my whole consciousness including all of its sensations. I rolled unto the floor beside the bed. While this happened, I did not feel bodiless, but I felt as if my body consisted of a substance constituted of a mixture of gas and liquid. To the present day I have never forgotten the combination of amazement and great surprise which gripped me when I felt myself falling onto the floor but the expected hard bounce never happened. Actually, had the movement unfolded in my normal body, my head would have had to collide with the edge of my bedside table. Lying on

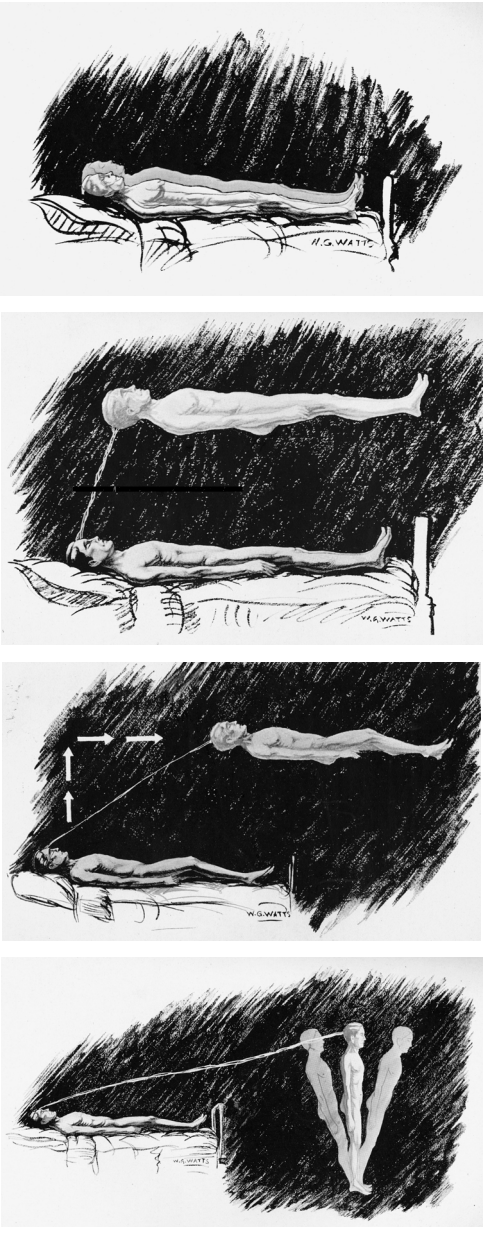

Figure 1: Kinematics of the PSM during OBE onset according to the classical Muldoon scheme. Reproduced from Muldoon and Carrington (1929).

> the floor, I was seized by terrible fear and panic. I knew that I possessed a body, and I only had one great desire – to be able to control it again. With a sudden jolt I regained control, without knowing how I managed to get back to it.

The prevalence of OBEs ranges from 10% in the general population to 25% in students, with extremely high incidences in particular subpopulations like, to take just one example, 42% in schizophrenics (Blackmore 1986; for an overview and further references see Alvarado 1986, 2000, p. 18p, and Irwin 1985, p. 174p). However, it would be false to assume that OBEs typically occur in people suffering from severe psychiatric disorders or neurological deficits. Quite the contrary, most OBE-reports come from ordinary people in everyday life situations. Let us therefore stay with non-pathological situations and look at another paradigmatic example, again reported by the Swiss biochemist Waelti (1983, p. 25; English translation TM):

> In a dazed state I went to bed at 11 p.m. and tried to fall asleep. I was restless and turned over frequently, causing my wife to grumble briefly. Now I forced myself to lie in bed motionless. For a while I dozed before feeling the need to pull up my hands, which were lying on the blanket, in order to bring them into a more comfortable position. In the same instant I realized that I was absolutely unable to move and that my body was lying there in some kind of paralysis. Simultaneously I could pull my hands out of my physical hands, as if the latter were just a stiff pair of gloves. The process of detachment started at the fingertips, in a way that could be clearly felt, almost with a perceptible sound, a kind of crackling. It was precisely the

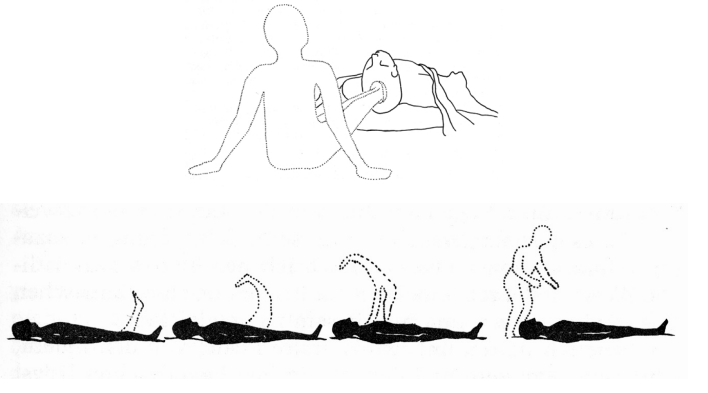

Figure 2: Kinematics of the phenomenal body-image during OBE onset. Two alternative, but equally characteristic motion patterns as described by the Swiss biochemist Waelti (1983).

> movement which I actually intended to carry out with my physical hands. With this movement, I detached from my body and floated out of it with the head leading. I gained an upright position, as if I was now almost weightless. Nevertheless I had a body consisting of real limbs. You have certainly seen how elegantly a jellyfish moves through the water. I could now move around with the same ease. I lay down horizontally in the air and floated across the bed, like a swimmer, who has pushed himself from the edge of a swimming pool. A delightful feeling of liberation arose within me. But soon I was seized by the ancient fear common to all living creatures, the fear of losing my physical body. It sufficed to drive me back into my body.

Sleep paralysis is not a necessary precondition for OBEs. They frequently occur during extreme sports, for instance in high-altitude climbers or marathon runners (Alvarado 2000, p. 184):

> A Scottish woman wrote that, when she was 32 years old, she had an OBE while training for a marathon. "After running approximately 12–13 miles ... I started to feel as if I wasn't looking through my eyes but from somewhere else. ... I felt as if something was leaving my body, and although I was still running along looking at the scenery, I was looking at myself running as well. My 'soul' or whatever, was floating somewhere above my body high enough up to see the tops of the trees and the small hills."

The classic OBE contains two self-models, one visually represented from an external perspective and one forming the center of the phenomenal world from which the first-person perspective originates. The representational integration of both components into one single conscious experience is achieved by the PMIR as introduced above. The representational and functional analysis of OBEs is difficult and challenging since there are many related phenomena, for example autoscopic phenomena during epileptic seizures in which only the first criterion is fulfilled.[2] Devinsky *et al.* (1998, p. 1080) differentiated between autoscopy in the form of a complex hallucinatory perception of one's own body as being external with "the subject's consciousness ... usually perceived within his body" and a second type, the classic OBE, including the feeling of leaving one's body and viewing it from another vantage point. The incidence of autoscopic seizures is possibly higher than previously recognized, Devinsky *et al.* (1998, p. 1085) found a 6.3% incidence in their patient population. Here is one of their case studies, demonstrating how OBEs can also develop from untypical etiologies like epileptic seizures (Devinsky *et al.* 1998, p. 1082):

[2] For a neurological categorization see Brugger *et al.* (1997), for an analysis focusing on the relevance of different degrees of body-centeredness in spatial perspective taking see Brugger (2002).

> A 29-year-old woman has had absence seizures since the age of 12 years. The seizures occur five times a week without warning. They consist of a blank stare and brief interruption of ongoing behavior, sometimes with blinking. She had an autoscopic experience at age 19 years during the only generalized tonoclonic seizure she has ever had. While working in a department store she suddenly fell, and she said, "... the next thing I knew I was floating just below the ceiling. I could see myself lying there. I wasn't scared; it was too interesting. I saw myself jerking and overheard my boss telling someone to 'punch the timecard out' and that she was going with me to the hospital. Next thing, I was in space and could see Earth. I felt a hand on my left shoulder, and when I went to turn around, I couldn't. Then I looked down and I had no legs; I just saw stars. I stayed there for a while until some inner voice told me to go back to the body. I didn't want to go because it was gorgeous up there, it was warm – not like heat, but security. Next thing, I woke up in the emergency room." No abnormalities were found on the neurological examination. Skull CT scan was normal. The EEG demonstrated generalized bursts of 3/s spike-and-wave discharges.

Seizures involving no motor symptoms or loss of consciousness and not being recognized by the patient may actually be more frequent than commonly thought.[3] One important feature of OBEs is that the phenomenal representation of the perceiving, acting self is confabulatory, while the representation of the remaining physical body from an external perspective is generally accurate. For instance, OBEs during seizures often clearly depict convulsive movements and automatisms very accurately, from a viewpoint above the body.[4] For many people who have actually lived through these phenomenal states this is an argument against the possibility of their hallucinatory nature.

However, it has to be noted that in the second self-model forming the object-component of the consciously modeled subject-object relationship, veridical content and confabulatory content are frequently *integrated* into a single whole. One patient noted that his body, perceived from an external perspective, was dressed in the same clothes he was wearing, but curiously he always had combed hair even when he knew his hair was uncombed before the onset of the episode (case 4; Devinsky *et al.* 1998, p. 1081). Another telling phenomenological difference is that some pa-

[3]For a case study of a patient who first experienced OBEs for a number of years and only later suffered from generalized seizures see Vuilleumier *et al.* (1997, p. 116).

[4]Devinsky *et al.* (1989, p. 1086) write: "Patient 39 was 'up there looking at myself convulsing, and my mother and the maid were screaming... I felt so sorry for them and my body.' Patient 40 watched her convulsive seizure, 'like being in a balcony', and observed the nurses placing a tongue depressor on her tongue and putting up the sides of the bed. Patient 33, who witnessed her complex partial seizure, clearly saw herself looking 'anxious, pale, and rubbing my hands, running aimlessly from one place to another'."

tients visually experience their body seen from above as not transparent and actually casting a shadow (cf. case 4 again). For some patients the double will be transparent, but slightly smaller than life size (case 9; Devinsky *et al.* 1998, p. 1082), while for other patients the seen body appears solid, but does *not* cast a shadow (case 2; Devinsky *et al.* 1998, p. 1081).

It may be relevant that even in spontaneous OBEs, clearly occurring in non-pathological contexts, the non-veridical or self-contradictory nature of particular forms experiential content may very well be cognitively available, not only after, but *during* the experience. Remember the report by Waelti (1983, p. 18) quoted above: "Actually, had the movement unfolded in my normal body, my head would have had to collide with the edge of my bedside table." Phenomenal kinesthetics and the underlying spatial frame of reference seem to be slightly dissociated in this case. This very fact itself is in turn available for cognitive processing, and for the formation of autobiographical memory.

As Alvarado (1997, p. 16) remarks, little systematic work has been conducted about the phenomenology of the experience (see also Alvarado 1986, 2000, p. 186). The content of OBEs certainly is globally available for attention and cognitive access. *Volitional* availability, however, is a highly variable component of the experience.[5] Many OBEs are dominated by a sense of passively floating. The two self-models that are active during an OBE are embedded into a coherent global state, into a single multi-modal scene forming an integrated model of reality. They are also activated within a window of presence, that is, the experience has no phenomenological characteristics of recollection or future planning – an OBE is something that is happening *now*. In fact, a considerable subset of OBEs is accompanied by the subjective experience of "hyperpresence" or "hyperrealism", particularly in those cases where a blending into or additional episode of religious ecstasy are reported. The phenomenal reality as modeled in the OBE certainly is a convolved and dynamic reality (see Metzinger 2003a, sections 3.2.4 and 3.2.5). OBEs are also first-person states: They clearly unfold under a single and unified first-person perspective generated by a PMIR. What makes them unique is that the object-component of the PMIR is formed by a self-model which is not a *subject*-model. You see your own body, and you recognize it as your own, but presently it is not the body *as subject*, the body as the locus of knowledge and of lived, conscious experience.

Of course, numerous exceptions exist in the colorful reports and the folklore about this kind of bodily self-consciousness. But the conceptually most interesting feature of OBEs arguably is that they are accompanied

[5]For an overview of the phenomenology see Irwin (1985, pp. 76ff). For an analysis of different case studies cf. Blackmore (1982a, pp. 56ff). For further references see Alvarado (2000).

by situations in which the subject- as well as the object-component of a phenomenal model of the subject-object-relationship is taken by a model of the self: *you* see your *own* body lying on the bed below you. Interestingly, this does not lead to a multi- or decentered overall state of consciousness. Only one of the currently active self-models functions as the "locus of identification". Typically, it is only the ethereal double hovering above, which is represented as the attentional subject, as the currently thinking self, and as the agent deliberately moving through space (see the marathon-runner example for an exception). It is interesting to note how OBEs, phenomenologically, are *not* states of disembodiment. On the contrary, there always seems to be a spatially located phenomenal self, even if its embodiment is reduced to a pure spatial point of visuo-attentional agency.

In general it seems safe to say that prototypical OBEs are fully transparent states. The model of reality generated during the experience is not experienced *as* a model, although in experienced subjects and practitioners this fact may well be *cognitively* available during the episode. It is precisely the transparency of OBEs, which has led generations of experiencers and theoreticians in many cultures and for many centuries in the past to naive-realistic interpretations of this deviant form of phenomenal self-modeling. However, many OBE subjects also report a "dreamlike quality, as if being awake in a dream". Among general dream variables like the prevalence of flying dreams, vividness, dream recall etc., the occurrence of lucid dreams is the most consistent predictor of OBEs (Alvarado 2000, p. 194p; see also section 7.2.5 in Metzinger 2003a). Blackmore (1986) found that subjects reporting deliberate, as compared with spontaneous, OBEs have a better ability to control and terminate dream content and more frequent flying dreams. An important hypothesis, yet to be empirically substantiated, therefore is that OBEs are just an additionally constrained subset of lucid dreams (see Blackmore 1982b and Metzinger 2003a, section 7.2.5).

In short, one may predict that a more systematic approach to the phenomenology of OBEs will yield different degrees of global transparency and opacity accompanying the experience. Moreover, the interrelatedness of this feature with other high-level variables should be investigated. For instance, OBEs can be functionally characterized as offline-activated states, because they typically occur when the body is asleep, paralyzed after an accident or under anesthesia. In these situations, globally available somatosensory input will be minimal. The PSM loses an important source of presentational content, driving and functionally anchoring it in internal stimulus sources under normal circumstances. Irwin (1985, pp. 308) has presented a theory of OBEs in which the notion of being "out of touch with somatic processes" plays a decisive role, either in terms of functional loss of input or in terms of attentional unavailability through habituation. An

interesting question, finally, is if OBEs satisfy the adaptivity-constraint: Can there be a *teleofunctionalist* analysis of OBEs? What function could this type of experience have for the organism as a whole? Here is a speculative proposal by Devinsky *et al.* (1998, p. 1088):

> There are several possible benefits that dissociative phenomena, such as autoscopy, may confer. For example, when a prey is likely to be caught by its predator, feigning death may be of survival value. Also, accounts from survivors of near-death experiences in combat or mountaineering suggest that the mental clarity associated with dissociation may allow subjects to perform remarkable rescue maneuvers that might not otherwise be possible. Therefore, dissociation may be a neural mechanism that allows one to remain calm in the midst of near-death trauma.

It is not at all inconceivable that there are physically or emotionally stressful situations, in which an information-processing system is forced to introduce a "representational division of labor" by distributing different representational functions into two or more distinct self-models (in what was previously called "multiple personality disorder", see Metzinger 2003a, section 7.2.4). The OBE may be an instance of transient functional modularization, of a purposeful separation of levels of representational content in the PSM.

For instance, if cut off from somatosensory input, or if flooded with stressful signals and information threatening the overall integrity of the self-model as such, it may be advantageous to integrate the ongoing conscious representation of higher cognitive functions like attention, conceptual thought and volitional selection processes into a *separate* model of the self. This may allow for a high degree of integrated processing, that is, for "mental clarity," by functionally encapsulating and thereby *modularizing* different functions like proprioception or attention and cognition in order to preserve at least some of these functions in a life-threatening situation. Almost all necessary system-related information is still glob-ally available, and higher-order processes like attention and cognition can still operate on it as it is presented in an integrated manner. But its distribution across specific subregions in phenomenal space as a whole has now dramatically changed. Only one of the two self-models is truly "situated" in the overall scene, integrated into an internally simulated behavioral space. Only one of them is immediately embodied and virtually self-present in the sense described. As it is fully transparent, it is a full-blown phenomenal self instantiating the phenomenal property of selfhood *for* the system. Frequently, both self-models integrated within a single OBE are constituted by spatial as well as non-spatial mental content.

Interestingly, the bodily self-model forming the object-component in this type of first-person experience never changes much in its spatial prop-

erties. The physical body viewed from an external perspective is very rarely distorted or changed in shape and size. However, the subject-component of the intentionality-relation modeled in these states may vary greatly (note how just the opposite principle holds for ordinary waking states). Some OBE subjects see or feel themselves in a weightless replica of their original body, others experience themselves as being in no body at all or in a kind of indeterminate form, such as a ball of light or an energy pattern (Alvarado 1997, p. 18; Green 1968) or even as "pure consciousness" (Alvarado 2000, p. 186).

This may indicate that spatial content is not strictly necessary in realizing the function fulfilled by the second self-model for the system as a whole. In other words, those higher functions such as attention, cognition and agency, which are integrated by the "dissociated" self, now are only *weakly embodied* functions. In order to be carried out they do not need the integration into a spatially characterized, explicit body image. Arguably, attentional and cognitive agency can *functionally* be decoupled from the process of autonomic self-regulation and the spatial self-representation necessary for generating motor behavior. Conceptually, this is an important insight about the human mind. As it is plausible to assume that also non-cognitive creatures like animals could undergo the type of fully disembodied OBEs described above, we may conclude that attentional agency actually is one of the essential core properties underlying the conscious experience of selfhood. Spatial self-representation and cognitive self-reference are not necessary for selfhood.

However, the prototypical OBE clearly takes place in an egocentric frame of reference possessing a spatial, bodily self-model as its origin. In this context, it may also be interesting to note that certain technological setups in virtual reality (VR) experiments – so-called second-person VR and telepresence systems (Heeter 1992, p. 264; see also Metzinger 2003a, section 8.1) – seem to achieve precisely the same effect, by creating the conscious experience of viewing one's own body as embedded into and interacting with a virtual world or the experience that there is a "real you" not currently inhabiting your body. Such technical systems offer an additional functional module (a graphic image or a robot body) through which subjects can control their own behavior. Participants in VR experiments of this type frequently describe their phenomenology simply as *being* an out-of-body experience, even if they have never had a natural OBE before (Heeter 1992). If it could be empirically confirmed that the spatiality of the attentional and cognitive self-model hovering above the self-as-object-component in the OBE-model of reality is not a strictly necessary condition, this would support the functional modularization hypothesis proposed here.

It is surprising to see how theoreticians exploring virtual environments today not only employ phenomenological notions like "presence" or "sit-

uatedness", but have already coined a terminological notion for what, under the self-model theory of subjectivity, would be the spatial partition of the PSM modeling motor properties of the organism: the "virtual body" (Barfield *et al.* 1995, p. 505). A virtual body is a part of an extended virtual environment, a dynamic and high-dimensional tool that can be used to control a robot at a distance, employing the virtual body as an interface. However, these authors also point out how the issue of "identification" is crucial in the context of teleoperator systems controlling distant robots, and how users of a virtual environment may actually reject their virtual body – just as some neuropsychological patients do (Barfield *et al.* 1995, p. 506). Most illustrative, however, is the notion of a "slave robot": To achieve telepresence, an operator has to rely on a high correlation between his own movements as sensed "directly" and the actions of the slave robot; and he ideally has to achieve an identification between his own body and that of the slave robot.

A virtual body, like a PSM, is an advanced interface to functionally appropriate and control a body. Virtual body and physical body may be separated by thousands of miles, and the interface used will (hopefully) only be episodically transparent. In the PSM-case, Mother Nature has solved all major interface problems millions of years ago, including a virtual body and extensive internal user modeling: Target system and simulating system are identical; and conscious subjectivity is the case in which a single organism has learned to enslave itself. This does not turn the system into a slave robot, but into an increasingly autonomous agent. Autonomy is conscious self-control, and an OBE is a situation in which self-control has been divided into different functional modules.

3.2 Psychology and the Functional Profile of OBEs

From a systematic point of view, thorough analyses of deviant phenomenal models of the self are of highest relevance for their psychological characterization. However, the quantity and quality of available scientific research is low. It is particularly low for OBEs, and also for neurophenomenological state-classes of related interest such as dissociative identity disorders (DID) or lucid dreams. It is hard to find empirical work that lives up to the methodological or conceptual standards of current cognitive neuroscience or analytical philosophy of mind.[6] Notable exceptions are to be found in the work by Irwin, Palmer and Blackmore.

[6]In this regard, the most important publications are Blackmore (1982a), Irwin (1985), and Palmer (1978). An excellent recent review is Alvarado (2000). A short overview concerning the literature and trends in research from the 19th century to 1987 can be found in Alvarado (1989), a review of modern developments from 1960 to 1984 concerning research on spontaneous out-of-body-experiences is Alvarado (1986). A review of three historical phases of psychological research since the 19th century can be found in Alvarado (1992). A more systematic overview concerning the phenomenology of OBEs can be found in Irwin (1985, pp. 76ff). Further discussion and a

Irwin proposes a model involving a shift in attentional processing during episodes of weakened somatosensory input and a kinesthetic completion of the somaesthetic body image mediated by a visual model of the environment, constructed from memory sources (Irwin 1985, pp. 306). As somaesthetic input is lost, other presentational subformats – like vision and kinesthesia – become more dominant and take its role in stabilizing the PSM. As Alvarado (2000, p. 203) points out, Irwin's model has received support from studies relating absorption and visuospatial abilities to the OBE and positively correlating synaesthesia-like items from a specific absorption scale to OBE frequency.

Palmer analyses OBEs as compensatory processes after events threaten the integrity of the overall self-model by causing fundamental changes in the body schema (see Palmer 1978). For Palmer, OBEs are just one of many routes the system can take to rescue its threatened phenomenal identity, to preserve the overall coherence of the self-model. As Alvarado (2000, p. 202) puts it, in Palmer's view the "OBE, then, is an attempt to prevent the jeopardy to one's identity from reaching awareness and precipitating a crisis."

Blackmore, to whom I am grateful for many exceptionally stimulating discussions, explicitly employs the concept of a "model of reality". Explicitly operating under the information-processing approach and analyzing the representational needs and resources of persons undergoing OBEs, she arrives at a theory describing OBEs as episodic models of reality, constructed by brains cut off from sensory input during stressful situations and having to fall back to internal sources of information. For instance, she draws attention to the remarkable fact that visual cognitive maps reconstructed from memory are organized from a bird's eye perspective in the majority of subjects and predicted that these persons would be more prone to having OBEs (see, for example, Blackmore 1982a, pp. 164; 1987). She also points out an important phenomenological feature of intended bodily *motion* in the OBE-state: frequently, the way in which OBE subjects move around in the currently active model of reality is not smooth, as in walking or flying, but occurs in discrete jumps from one salient point in the cognitive map to the next. Blackmore's observation emphasizes that, whatever else OBEs are, they certainly are internally simulated behavioral spaces. This phenomenological observation indicates that frequently these behavioral spaces, typically simulated by a brain under great stress, are *spatially underdetermined* – i.e., they are coarse-grained internal sim-

review of attempts towards the development of empirical taxonomies and typologies of the OBE are given in Alvarado (1997). Blackmore (1982a, pp. 56ff) offers an analysis of different case studies. Reports of OBEs in non-Western cultures and of previous scientific studies can be found in Blackmore (1982a, pp. 71ff, pp. 82ff). Wolfradt and Watzke (1999) present an interesting recent study of the relationship between depersonalization, schizotypal personality traits, and deliberate OBEs.

ulations of landmarks and salient spots in certain perceptual scenes that were seen and acted upon at an earlier stage in life. The general idea in Blackmore's theory is that OBEs are transparent phenomenal simulations of the world, which are highly realistic because they include a partially veridical representation of a phenomenal body and are organized from an external "third-person" visual perspective (Blackmore 1984, 1987).

All these approaches are consistent with the self-model theory of subjectivity. They are explicitly presented as *psychological* theories and do not assume any non-physical carrier substance for conscious experience being in existence or actually leaving the body during an OBE. They are parsimonious by being simulational, and not representational, theories of the OBE; because they do not assume that there is an actual *representatum* in the environment of the physical body, corresponding to the PSM as an exteriorized second entity. However, taking a more careful look at abstract, non-spatial aspects of the phenomenal self in these states, one discovers how the subject-component of the PMIR in the OBE-state is not completely empty. An attentional and cognitive subject engaging in selective processing is modeled, and actually *in existence*: OBE subjects generally have good control over their attentional and their thought processes as such, even if almost all the *contents* of these processes may be hallucinatory.

3.3 Neural Correlates of OBEs

Is there a neural correlate of the out-of-body experience? Is there a minimal set of brain properties, which is sufficient for bringing about an OBE (Metzinger 2000)? It was recently shown that phenomenal states closely resembling OBEs can be induced by electrically stimulating the brain at the right angular gyrus. This leads to the empirical hypothesis that a disintegration of somatosensory and vestibular information may be an important factor in generating an OBE.[7] These results created renewed interest in OBEs and put them into the context of recent work on body illusions and, more generally, self-processing (e.g., Gillihan and Farah 2005).

It has long been known that OBEs not only occur in healthy subjects, but in certain clinical populations (e.g., epileptic patients) as well. In a recent study, Blanke *et al.* (2004) could localize the relevant brain lesion or dysfunction at the temporo-parietal junction (TPJ; see Fig. 3) in five out of six patients. They argued that two separate pathological conditions may be necessary to cause an OBE: (1) a disintegration in the

[7] See Blanke *et al.* (2002). For a more detailed hypothesis concerning the role of the temporo-parietal junction see Blanke *et al.* (2004), cf. also Blanke *et al.* (2005) and Tong (2003). For recent reviews see Blanke and Arzy (2005) and Bünning and Blanke (2005).

self-model or "personal space" (brought about by a failure to integrate proprioceptive, tactile, and visual information regarding one's own body) and (2) an additional disintegration between external, "extrapersonal" visual space and the internal frame of reference created by vestibular information. The experience of seeing one's own body in a position that does not coincide with its felt position could therefore be caused by cerebral dysfunction at the TPJ. Both types of functional disintegration lead to the representational configuration described above. The proposal of a twofold disintegration extends previous models about phantom limbs and supernumerary phantom limbs.

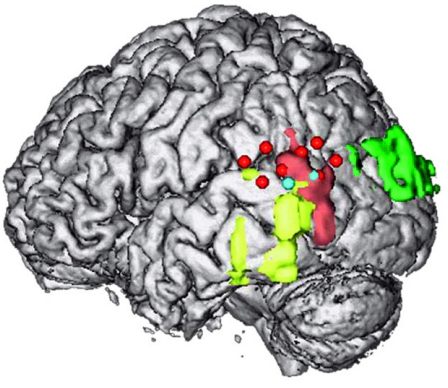

Figure 3: Results of a mean lesion overlap analysis of MRI scans in five patients with OBEs or related experiences. The analysis centers on the temporo-parietal junction. Blue points show the locus of electrical cortical stimulation artificially inducing OBE-like phenomenal states. Colored areas are activated during epileptic seizures. Reproduced with permission from Blanke *et al.* (2004).

Figure 4 shows MRI data from an epileptic patient with electrodes implanted in the left hemisphere. Again, significant activation at the TPJ was observed (Blanke *et al.* 2005). Using evoked potential mapping, it was additionally shown how a selective activation of the TPJ takes place 330–400 ms after healthy volunteers mentally imagined themselves in a position and visual perspective characteristic of an OBE. They also demonstrated that it is possible to impair this mental transformation of the bodily self-model by interfering with transcranial magnetic stimulation at this specific location. In an epileptic patient with OBEs caused by damage at the TPJ it could be shown that mimicking the OBE-PSM (i.e., by mentally simulating an OBE like the ones experienced before)

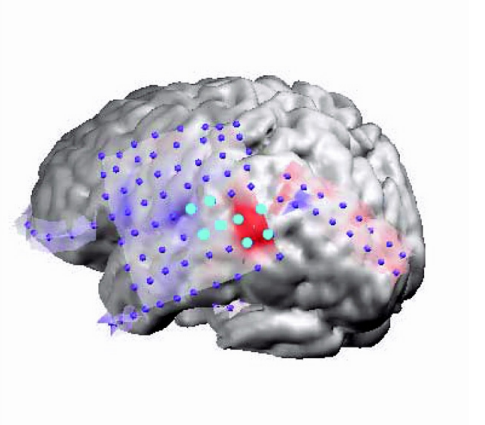

Figure 4: MRI of an epileptic patient with implanted electrodes overlying the lateral convexity of the left brain hemisphere. The epileptic focus, for which the discharge induced an OBE, is indicated by eight turquoise electrodes at the TPJ. The color shaded area shows the electrodes where the strongest evoked potentials were observed when the patient performed a paradigm mimicking the OBE-PSM. Reproduced with permission from Blanke *et al.* (2005).

the seizure focus was partially activated (Blanke *et al.* 2005). These observations provide strong evidence for an anatomical link between three similar types of phenomenal mental content: (i) seizure-caused OBEs, (ii) intended mental simulation of an OBE (i.e., imagery) in healthy subjects, and (iii) intended mental simulation of an OBE (i.e., imagery) in epileptic patients.

Let me explicitly draw the attention of those readers interested in empirical details to the references given above. Interim conclusions are that the TPJ is a locus where low-level processing of vestibular information and high-level processes like the generation of a unified phenomenal self, the phenomenal property of agency and the emergence of a multimodal first-person perspective (i.e., a consciously experienced egocentric frame of reference) critically interact.

The empirical material briefly discussed here demonstrates how much can be learned about the mechanisms underlying human self-consciousness in *normal* cases, about those functional and representational properties realized by the human brain that support the generation of a phenomenal self-model in *non-pathological* situations. Clinical neurology is capable of contributing substantially to an empirically plausible multi-level theory of human consciousness. In this project, modern research on OBEs is particularly relevant, because it helps unveiling the fine-grained functional architecture underlying the conscious self-model of human beings. The self-model theory of subjectivity represents an attempt to provide a wide and coherent conceptual framework into which these data fit.

4. Metatheoretical Conclusions and Conceptual Implications

First, let us not jump to false conclusions. The OBE may turn out not to be a distinct and unified target phenomenon, and it may possess a variance across populations. For instance, the initial "exit phenomenology" (i.e., the first seconds of an OBE) clearly seems to differ between spontaneous OBEs in healthy subjects and in clinical populations. It may yet be different again in followers of particular spiritual practices, where the occurrence is not spontaneous but the subjects are healthy. Second, there could be a significant neurophenomenological overlap between lucid dreaming and OBEs, and between other types of neurological disorder like autoscopy (Brugger 2002), the supernumerary phantom limb phenomenon (Ramachandran and Hirstein 1998), or body illusions in general. Third, it is important to be clear about potential *ontological* conclusions from the material presented in this paper. Even if a reductive explanation of all types of OBEs as deviant configurations of the human PSM should be achieved in the future, and even if the hypothesis about the history of the *concept* of a soul presented here is correct, it still remains logically possible that souls do exist. We would then not need the concept of a soul any more for the purposes of science or philosophy, because it would not play an explanatory role in any rational, data-driven theory any more. We would also have a deeper understanding of its genesis in human culture. But from a strictly logical point of view it remains possible that one day we discover a sense in which it is not an empty concept at all.

The present renaissance of rigorous research on the OBE allows us to see how the phenomenal content of OBEs can successfully be redescribed at a *representationalist* level of analysis. OBEs can be analyzed as a special form of mental self-representation or deviant self-modeling (Metzinger 2003a). We are beginning to understand the more fine-grained functional properties underlying this process of deviant self-modeling, the sites of their neural implementation, and the systematic effect of their disturbance on the phenomenology of human subjects and patients. This, in turn, helps to develop a comprehensive, unified theory of the PSM, the conscious human self-model – a theory of self-consciousness which is both conceptually coherent and empirically plausible. As a matter of fact, the current development may even be seen as historic: Cognitive neuroscience now clearly starts to make substantial contributions to ancient philosophical projects such as the ideal of self-knowledge.

From a philosophical perspective, OBEs are interesting for a number of reasons. First, in the purely systematic context of a representational theory of mind, they provide us with a unique phenomenal configuration: OBEs are global, phenomenal models of reality in which two self-models, but only one first-person perspective exist. That is, we have a more

or less stable, centered model of reality that contains a PMIR. The interesting point is that during some episodes the subject- as well as the object-component of the transparent model of the intentionality-relation is constituted by a representational structure actually purporting to depict the experiencing person *herself*. OBEs show that self-models are not necessarily *subject*-models: You can represent something as your *own* body, without representing it as an agent with which you are identical, and you can do so under a *perceptual* model of the subject-object-relation. OBEs are like a "perceptualized" variant of reflexive self-consciousness. OBEs constitute a strong argument for the thesis that, while an accompanying bodily self-model may be fully "confabulated" by subpersonal mechanisms fighting for global coherence, the phenomenal locus of the self is always where the locus of cognitive and attentional *agency* is (see section 3.1). Interestingly, this is not true of *bodily* agency (recall the example of the marathon runner above). It is easy to conceive of systems that are not cognitive, but only attentional agents (for instance, animals) and nevertheless have OBEs. Therefore, the experience of attentional agency may be the core of phenomenal selfhood and perspectivalness and the origin of all consciously experienced intentionality.

More generally, the phenomenological concept of an OBE seems to be a cluster concept, and the phenomenal state-class picked out by this concept is characterized by a high degree of variability in phenomenal content. However, there are a number of further and essential features. In whatever way the ethereal "double" or "doppelgänger" leaving the physical body is phenomenally modeled, the cognitive and attentional subject – the self-model modeling the system as a cognitive and attentional agent (see Metzinger 2003a, sections 6.4.3 and 6.4.4) – always forms the phenomenal "locus of identity". This locus is invariably represented as the subject-component of the represented subject-object-relationship, thereby generating the structural feature of the overall model of reality referred to as perspectivalness. There are higher-order types of self-consciousness (i.e., self-models internally satisfying the perspectivalness constraint, see Metzinger 2003a, section 6.4.4; Metzinger 2005), where the PMIR points from a second-order self-representation to a first-order self-representation – as in phenomenologically *inward*-directed attention and self-related cognition. OBEs are unique in being simulations of *perceptual* PMIRs, establishing a system-system relationship modeled within a spatial frame of reference. It is as if, in situations where the self-model can no longer be anchored in internal somatosensory input or a low-level egocentric frame of reference (see Metzinger 2003a, section 5.4), higher cognitive functions like attentional processing or categorical thought simply take over in *centering* the global model of reality. In this way some persons undergoing an OBE truly are disembodied, thinking selves in a neurophenomenologically reduced version of the original Cartesian sense. However, it is not subjec-

tively available to them that all this is just a *model* of reality generated by their central nervous system.

This again leads to a number of issues of a more general philosophical interest. For anyone who actually had that type of experience it is almost impossible not to become an ontological dualist afterwards.[8] In all their realism, cognitive clarity and general coherence, these phenomenal experiences almost inevitably lead the experiencing subject to conclude that conscious experience can, as a matter of fact, take place *independently* of the brain and the body: what is phenomenally possible in such a clear and vivid manner must also be metaphysically possible or actually the case. Although many OBE reports are certainly colored by the interpretational schemes offered by the metaphysical ideologies available to experiencing subjects in their time and culture, the experiences as such must be taken seriously. Although their conceptual and ontological interpretations are often seriously misguided, the truthfulness of centuries of reports about ecstatic states, soul-travel and second bodies as such can hardly be doubted.

From an open-minded, rational, and metatheoretical perspective OBEs are not only a problem for philosophy of mind and phenomenology, but for *epistemology* as well. What about persistent claims regarding veridical perception during OBE states? I think that the failure of previous research on this issue has shown that we must find another way to solve the problem. The research strategy I propose is to proceed with a fined-grained representational and functional analysis of OBEs and related phenomena until we are able to make the target phenomenon *repeatable*, an object of investigation that can be reliably reproduced in a rigorously controlled experimental setting. Then we can directly investigate claims with regard to extrasensory perception during the OBE state in a systematic manner. In this context, let me point to a logical possibility, which is rarely noticed: OBEs may, at the same time, be both confabulatory states or complex hallucinations *and* information-bearing states correctly representing certain aspects of the environment. In some cases, an OBE may simply be the way in which the brain searches for a maximally coherent global state: It may be the way in which a human brain desperately tries to explain to itself that it *has* some information, namely in a situation where the causal history of successfully extracting this information from the environment cannot be understood or represented.

At this point it is interesting to note how all conscious models of re-

[8]For instance, 73% of respondents to a study by Osis (1979) claimed a revised attitude about life after death after experiencing an OBE, 67% reported a reduction in their fear of death, and 66% in a study by Gabbard and Twemlow (1984) claimed to have actually adopted a belief in life after death. For further references see Alvarado (2000, p. 188). For a recent empirical study of near-death experiences in cardiac arrest survivors see Parnia *et al.* (2001).

ality and the self in it can be read as ontologies and as epistemological metaphors. As phenomenal ontologies they are non-propositional "theories" – internal, neurobiologically realized models – about what actually *exists* from the brain's point of view. As epistemological metaphors they are theories about how the organism actually comes to *know* about the existence of this reality. Under a naive-realistic interpretation they can then become *theoretical* ontologies – folk-phenomenology turns into folk-metaphysics, as it were. I propose that this happened in the historical transition from truthful, first-person phenomenological reports about OBEs to the proto-concept of mind.

In conclusion, first-person reports about OBEs are available in abundance not only from all times, but also from many different cultures. There is a culturally invariant core to the phenomenon which obviously forms a coherent cluster of properties. On the other hand, it is too early to decide whether or not the high internal correlation strength characterizing this set of phenomenal properties ultimately justifies to treat all OBEs as members of one single and distinct phenomenal state-class. Nevertheless, the *prima facie* assumption remains a rational and productive working hypothesis. In this sense, the experience of a soul-like entity, an ethereal or astral body leaving the physical body during sleep, after accidents and in death can be called a "phenomenological archetype" of mankind. Following this line of thought, three independent, but complementary conclusions can be drawn.

First, an out-of-body experience actually is a *neuro*phenomenological archetype: The functional core of this kind of phenomenal state is formed by a *culturally invariant neuropsychological potential* common to all human beings. Under certain conditions, the brains of *all* human beings, through specific properties of their functional and representational architecture, which have yet to be empirically investigated, allow for this set of phenomenal models of reality. A plausible and rational working hypothesis is that this set of models of reality may turn out to be a discrete set, forming an individual, clearly circumscribed goal for empirical research. Correspondingly, a distinct, minimally sufficient neural correlate for the OBE-state in humans is likely to exist, and, in principle, a functionalist analysis of the phenomenon can be developed from a more fine-grained representationalist analysis. Maybe, in some distant future, even machines can engage in soul-travel.

The notions of a PSM and a PMIR (see section 3.1., and Metzinger 2003a, 2005) may be assumed as a viable starting point for operationalizing OBEs. However, this assumption may be false, and it will also be important to find out how high the degree of cultural invariance in OBEs actually is. Maybe OBEs are *not* a distinct theoretical entity, but just a subcluster of prelucid dreams, body illusions or a tendency towards depersonalization, intuitive thinking, or certain schizotypal personality

traits (Wolfradt and Watzke 1999). In any case, the second point which makes OBEs an interesting target for philosophical analysis is that they are likely to form a *neuroanthropological constant*. Given the necessary neurofunctional configuration, this implies a potential for a certain type of experience shared by all human beings. Non-linguistic creatures not embedded into a rich cultural environment could have these experiences as well. However, OBEs could be *strong* first-person phenomena (in the sense of Baker 1998, as discussed in Metzinger 2003a, section 6.4.4; see also Metzinger 2003b) only in humans, namely by additionally being self-ascribed on a *conceptual* level. On our planet, so far only human beings have had OBEs *and* the capacity to think and communicate about them, because only they have had the necessary brain structures. Humans were the first beings capable of conceptually self-ascribing these experiences to themselves, culturally embedding them through folk-phenomenological discourse and the formation of a proto-concept of mind. Hence, the potential to have *strong* OBEs is proposed as a neuroanthropological constant.

The third important aspect which makes OBEs interesting from a philosophical history-of-ideas perspective – and which again highlights the relevance of rigorous, empirical research programs from a purely meta-theoretical perspective – has to do with the origins of *theoretical* self-awareness. My last proposal is that phenomenal states such as OBEs, which indicate a commonality in the neurofunctional architecture underlying the process of conscious human self-modeling, are the historical root of the proto-concept of mind. The proto-concept of mind eventually developed into Cartesian dualism and idealistic theories of consciousness. In short, the particular phenomenal content of OBEs led human beings to believe in a *soul*. Let us simply call this the "soul-hypothesis": After the evolution of brains had reached a stage at which OBEs in terms of strong, conceptually mediated forms of phenomenal self-modeling became possible, it was only natural – at a theoretical level – to assume that something like a soul actually exists. Given the epistemic resources of early mankind, it was a highly rational belief to assume the possibility of disembodied existence. And it was the PSM of *homo sapiens* which made this step possible.

The history of the concept of mind is a history of increasing differentiation and abstractness. Initially there was a theory of something concrete, an ethereal and spatially extended double, a breath of life. Eventually we find something entirely unworldly, an abstract, ideal principle. It is remarkable how the best theories of mind available today again turn it into a concrete process, fully endowed with temporal and spatial properties. However, in the light of contemporary cognitive neuroscience it is even more remarkable how, at the beginning of human theorizing about mind and consciousness, we find a very similar basic motive across very different cultural contexts: the idea of a "subtle body" which is independent

of the physical body and the true carrier of higher mental functions like attention and cognition (Mead 1919).

Historically, the dualist tradition in philosophy of mind is rooted in these early proto-theories. These theories, I propose, may in turn be motivated by naive-realistic interpretations of early first-person reports about OBEs. We saw already how many of the deviant models of reality and self characterizing altered states of consciousness and pathological neurophenomenological configurations may have a hidden heuristic potential, because they can also be read as metaphysical or epistemological metaphors. In a way, they are the brain's own philosophy. As phenomenal ontologies they are neurobiologically realized assumptions about what actually *exists* from the brain's point of view. Taken as an ontological metaphor, the phenomenology of OBEs inevitably leads to dualism, and to the concrete idea of an invisible, weightless, but spatially extended *second body*. This, then, may actually be the folk-phenomenological ancestor of the soul, and of the philosophical proto-concept of mind: The soul is the OBE-PSM.

Centuries of phenomenological reports describing it as a *subtle* body pointed in the right direction, and now we begin to see how it actually is a purely *informational* structure modeling bodily self-experience in cases of absent or disintegrated somatosensory/vestibular input. Therefore, in order to not only have an empirically grounded theory of conscious experience, but also to understand the neurofunctional and neurophenomenological underpinnings of the persisting intuition that such a theory leaves out something important, it will be of highest relevance to achieve a fuller understanding of this type of phenomenal experience. The sketched hypotheses of the *culturally invariant neuropsychological potential* (CINP), of the *neuroanthropological constant* (NAC), and of the *soul* may be good starting points to finally take phenomenology seriously. The traditional concept of an immortal soul, which can exist independently of the physical body, probably possesses a phylogenetically recent *neurophenomenological correlate*: the OBE-PSM, the type of deviant phenomenal self-modeling described in this contribution.

Acknowledgments

I wish to thank two reviewers, Olaf Blanke and Harald Atmanspacher for valuable discussions. I also wish to thank Sue Blackmore and Peter Brugger for critical comments on earlier versions of this paper; and Olaf Blanke and Ernst Waelti for the permission to reproduce their figures.

References

Alvarado C.S. (1986): Research on spontaneous out-of-body experiences: A review of modern developments, 1960–1984. In *Current Trends in PSI Research*, ed. by B. Shapin and L. Coly, Parapsychology Found., New York, pp. 140–174.

Alvarado C.S. (1989): Trends in the study of out-of-body-experiences: An overview of developments since the nineteenth century. *Journal of Scientific Exploration* **3**(1), 27–42.

Alvarado C.S. (1992): The psychological approach to out-of-body-experiences: A review of early and modern developments. *The Journal of Psychology* **12**, 237–250.

Alvarado C.S. (1997): Mapping the characteristics of out-of-body experiences. *Journal of the American Society for Psychical Research* **91**, 13–30.

Alvarado C.S. (2000): Out-of-body experiences. In *The Varieties of Anomalous Experience: Examining the Scientific Evidence*, ed. by E. Cardeña, S.J. Lynn, and S. Krippner, American Psychological Association, Washington D.C., pp. 183–218.

Baker L.R. (1998): The first-person perspective: A test for naturalism. *American Philosophical Quarterly* **35**, 327–346.

Blackmore S. (1982a): *Beyond the Body: An Investigation of Out-of-the-Body Experiences*, Granada, London.

Blackmore S. (1982b): Out-of-body experiences, lucid dreams, and imagery: two surveys. *The Journal of the American Society for Psychical Research* **4**, 301–317.

Blackmore S. (1984): A psychological theory of the out-of-body-experience. *Journal of Parapsychology* **48**, 201–218.

Blackmore S.J. (1986): Spontaneous and deliberate OBEs: A questionnaire survey. *Journal of the Society for Psychical Research* **53**, 218–224.

Blackmore S.J. (1987): Where am I? Perspectives in imagery and the out-of-body experience. *Journal of Mental Imagery* **11**, 53–66.

Blanke O., Ortigue S., Landis T., and Seeck M. (2002): Stimulating illusory own-body perceptions. *Nature* **419**, 269–270.

Blanke O., Landis T., Spinelli L., and Seeck M. (2004): Out-of-body experience and autoscopy of neurological origin. *Brain* **127**, 243–58.

Blanke O. and Arzy S. (2005): The out-of-body experience. Disturbed self-processing at the temporo-parietal junction. *The Neuroscientist* **11**, 16–24.

Blanke O., Mohr C., Michel C.M., Pascual-Leone A., Brugger P., Seeck M., Landis T., and Thut G. (2005): Linking out-of-body experience and self processing to mental own-body imagery and the temporoparietal junction. *Journal of Neuroscience* **25**, 550–557.

Brugger P. (2002): Reflective mirrors: Perspective-taking in autoscopic phenomena. *Cognitive Neuropsychiatry* **7**, 179–194.

Brugger P., Regard M., and Landis T. (1997): Illusory reduplication of one's own body: phenomenology and classification of autoscopic phenomena. *Cognitive Neuropsychiatry* **2**, 19–38.

Bünning S. and Blanke O. (2005): The out-of-body experience: Precipitating factors and neural correlates. *Progress in Brain Research*, in press.

Churchland P.M. (1981): Eliminative materialism and the propositional attitudes. *Journal of Philosophy* **78**, 67–90.

Devinsky O., Feldmann E., Burrowes K., and Bromfield E. (1989): Autoscopic phenomena with seizures. *Archives of Neurology* **46**, 1080–1088.

Gabbard G.O. and Twemlow S.W. (1984): *With the Eyes of the Mind: An Empirical Analysis of Out-of-Body States*, Praeger, New York.

Gillihan S.J. and Farah M.J. (2005): Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin* **131**, 76–97.

Green C. (1968): *Lucid Dreams*, Institute of Psychophysical Research, Oxford.

Heeter C. (1992): *Being* there: the subjective experience of presence. *Presence* **1**, 262–271.

Irwin H. (1985): *Flight of Mind*, Scarecrow Press, Methuen NJ.

Lishman J.R. and Lee D.N. (1973): The autonomy of visual kinaesthesis. *Perception* **2**, 287–294.

Mead E.R.S. (1919): *The Doctrine of the Subtle Body in Western Tradition*, John M. Watkins, London.

Metzinger T. (1993): *Subjekt und Selbstmodell*, Mentis, Paderborn (2nd edition 1999).

Metzinger T., ed. (2000): *Neural Correlates of Consciousness*, MIT Press, Cambridge.

Metzinger T. (2003a): *Being No One. The Self-Model Theory of Subjectivity*, MIT Press, Cambridge (2nd edition 2004).

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

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

Muldoon S. and Carrington H. (1929): *The Projection of the Astral Body*, Rider & Co., London.

Oeing-Hanoff L., Verbeke G., Schrott B., Nobis H.M., Marquard O., and Rothe K. (1974): Geist. In *Historisches Wörterbuch der Philosophie, Band 3*, ed. by J. Ritter and K. Gründer, Schwabe, Basel, pp. 154–204.

Osis K. (1979): Insider's view of the OBE: A questionnaire study. In *Research in Parapsychology*, ed. by W.G. Roll, Scarecrow Press, Methuen NJ, pp. 50–52.

Palmer J. (1978): The out-of-body-experience: A psychological theory. *Parapsychology Review* **9**, 19–22.

Parnia S., Waller D.G., Yeates R., and Fenwick P. (2001): A qualitative and quantitative study of the incidence, features and aetiology of near death experiences in cardiac arrest survivors. *Resuscitation* **48**, 149–156.

Ramachandran V.S. and Hirstein B. (1998): The perception of phantom limbs. The D.O. Hebb lecture. *Brain* **121**, 1603–1630.

Schrott B. (1974): Geist III. Der jüdische und christliche Geistbegriff. In *Historisches Wörterbuch der Philosophie, Band 3*, ed. by J. Ritter and K. Gründer, Schwabe, Basel, pp. 162–169.

Stanford R.G. (1976): A study of motivational arousal and self-concept in psi-mediated instrumental response. *Journal of the American Society for Psychical Research* **70**, 167–178.

Tong F. (2003): Out-of-body experiences: from Penfield to present. *Trends in Cognitive Sciences* **7**, 104–106.

Verbeke G. (1974): Geist II. Pneuma. In *Historisches Wörterbuch der Philosophie, Band 3*, ed. by J. Ritter and K. Gründer, Schwabe, Basel, pp. 154–162.

Vuilleumier P., Despland P.A., Assal G., and Regli F. (1997): Héautoscopie, exta-se et hallucinations expérientielles d'origine épileptique. *Revue Neurologique* **153**, 115–119.

Waelti E. (1983): *Der dritte Kreis des Wissens*, Ansata, Interlaken.

Wolfradt U. and Watzke S. (1999): Deliberate out-of-body experiences, depersonalization, schizotypal traits and thinking styles. *Journal of the American Society for Psychical Research* **93**, 249–257.

Received: 24 January 2005
Revised: 25 April 2005
Accepted: 27 April 2005

Reviewed by Eberhard Bauer and another, anonymous, referee.