

Sonja Rinofner-Kreidl, Graz

## The Limits of Representationalism

### A Phenomenological Critique of Thomas Metzinger's Self-Model Theory

#### **Abstract**

*Thomas Metzinger's self-model theory offers a framework for naturalizing subjective experiences, e.g. first-person perspective. These phenomena are explained by referring to representational contents which are said to be interrelated at diverse levels of consciousness and correlated with brain activities. The paper begins with a consideration on naturalism and anti-naturalism in order to roughly sketch the background of Metzinger's claim that his theory renders philosophical speculations on the mind unnecessary (I). In particular, Husserl's phenomenological conception of consciousness is refuted as uncritical and inadequate. It will be demonstrated that this critique is misguided. (II). The main deficiencies of Metzinger's theory will be elucidated by referring to the conception of phenomenal transparency (III) which will be compared to a phenomenological idea of transparency (IV). Then we shall enlarge our critical horizon by focusing on some implications of representationalism, including reification of consciousness, brain-Cartesianism, and exclusion of the social dimension (V). Finally, we shall take up our meta-theoretical reflections on the naturalism debate (VI).*

#### **I. Cognitive Science versus Phenomenology: Points of Departure**

Cognitive scientists usually take for granted that it is, on principle, possible to naturalize subjective experiences. There is much debate on the conditions and possibility of this enterprise. However, there is hardly any fundamental doubt concerning the feasibility and adequacy of naturalization projects. The idea of subjectivity involved in these theories mainly refers to *qualia* and diverse modes of self-reference. Accordingly, a considerable part of the debate on the prospects of naturalization programs in the field of cognitive science focuses on the relation between presentational and representational contents of mind, both being grasped in terms of multi-realizable functional states.

Phenomenologists insist on the practical dimension of subjectivity and make a point of questioning the methodical and conceptual presuppositions of cognitive scientific models of the mind. From this point of view, it is obvious that an effective refusal of *strong naturalization projects which ignore or distort our life-worldly experience* requires more than exploring the consistency and empirical plausibility of the theories in question. A phenomenological critique of such theories transcends the limits of an immanent critique, however elaborately this may be done. Therefore, the onus rests with the phenomenologists to show that their objections cannot be passed over lightly or ridiculed as dogmatically ignoring the efficiency of modern natural science.

The core of what phenomenologists have to say with regard to strong projects of naturalizing the human mind is this: Refuting naturalism is tantamount to appreciating the work of natural science whose impressive success is exactly enabled by the fact that natural science, due to its methodical idealizations and specific theoretical interest, is always an undertaking of limited scope (cf. Husserl 1962). Naturalism arises whenever this fact is obscured, displaced or explicitly denied. Neglecting the limited scope of scientific theories goes hand in hand with neglecting the function and status of the subject whose experiences are the starting-point of all scientific theorizing. It may therefore be said that self-forgetfulness is a significant mark of the natural scientific attitude (Husserl 1952a, pp. 183–184). It manifests itself in a strong affinity to a purely mechanical or technical approach to scientific methods. Heidegger's fundamental ontology, which claims to go beyond traditional distinctions such as subject/object or theory/practice *via* an existential analysis of being-human (*Dasein*), represents an alternative, phenomenological critique of naturalism. According to this project, the limited scope of scientific investigations is explained by referring to the *ontological difference*, i.e. the difference between Being (*Sein*) and particular spheres of entities (*Seiendes*), by querying the presuppositions implied in the theoretical attitudes of science and philosophy and, more specifically, by maintaining a foundational relation between a logical idea of science and an existential idea of science (Heidegger 2001, § 69b). If phenomenologists are right in arguing that the abilities and achievements of the subject (Husserl) or of being-human (Heidegger), genetically viewed, underlie all scientific investigations, then it goes without saying that what is at stake in the naturalism debate is nothing less than our idea of science. If the exclusion of the subject or of being-human is the most fundamental objection to strong and dogmatic types of naturalism, then it seems plausible that cognitive scientists aim at strengthening their position by demonstrating that subjective experiences can be naturalized. In reply to this, phenomenologists argue that constructions of a naturalized subjectivity, again, can be understood only by referring to a subjectivity both embedded in a life world and not naturalized. In the following we shall confine ourselves to a Husserlian-style, anti-naturalistic reasoning, for the naturalization project at issue explicitly refers to Husserl's idea of intentionality. The latter is given prominence as the most influential old-fashioned and unscientific model of the human mind which cognitive science is expected to overcome.

## II. Husserl and Metzinger: Two Incompatible Models of the Human Mind

Phenomenology is, basically, interested in phenomenality. It claims to grasp phenomena, i. e. things and processes, solely with respect to their appearance. For conceptual reasons, there is no phenomenon *in itself*. Phenomena are intrinsically related to some consciousness for whom they are presently given. According to our natural attitude, we do not encounter phenomena but those things we are directed at for multifarious reasons, for instance, in order to gain some knowledge about them, technically modify them or consume them. To become aware of phenomena means to become aware of the fact that there is an intentional structure lying beneath our ordinary way of handling things and looking at the world. This awareness requires a change of attitude which involves a reflective stance.

Husserl's transcendental phenomenology establishes an attitude towards attitudes (Husserl 1952a, pp. 174, 179). This higher order reflexivity can, among other things, be directed at scientific attitudes. In this case it must be shown that every theory is based on a specific correlation between a method and its object of investigation. For example, natural science always operates on the basis of a tacit agreement concerning what qualifies as *nature* or *natural object* (Husserl 1952a, p. 2; Husserl 2001, p. 10). Moreover, every theoretical attitude (including a philosophical one, cf. Husserl 1952a, p. 146) refers to a specific theorizing subject, the latter being a methodical fiction which must not be confused with a full-blooded human person. It goes beyond the specific theoretical interest of natural science to dig into these correlations of method, object and subject (cf. Husserl 1952a, pp. 287–288, 355). Phenomenology, on the other hand, is the endeavour to disclose the hidden structure of intentionality in a methodically disciplined manner. It makes explicit the correlation between consciousness and world with regard to a variety of intentional relations.

Reading works of Husserl, we constantly stumble on the term »consciousness«. This might be misleading. Phenomenology is not interested in investigating consciousness as an *object* of scientific, e.g. psychological, concern cutting off its relations to other regions of being. Consciousness is of *philosophical* interest only insofar as it functions as the medium of experiences embedded in a pre-given world. If we take consciousness in its pure intentional function, leaving aside the existential assumptions we normally attach to it with regard to both the experience and its object, we discover that »pure« consciousness incorporates all other regions of being: everything can be considered a phenomenon. Every physical or higher-order object or process can be considered as intentionally related to consciousness. According to Husserl's transcendental-phenomenological idealism, pure consciousness, which renders possible the appearance of any object whatsoever, cannot, on its turn, be conceived of as appearing in the mode of an object. The result would otherwise be an infinite regress of consciousness, since we would have to ask *for whom* this appearing consciousness-*phenomenon* were given and so forth ad infinitum.

Phenomenologically viewed, consciousness is the place where the world is brought to appearance. Consciousness, due to its physical foundation, always occupies a certain place within the world. On the other hand, the world is brought to appearance by consciousness *in terms of meaning constitution*. (A pure ego not embodied could not have any appearances.) Husserl's *paradox of subjectivity* refers to this double role of constituting world and being part of an already constituted world. This is far from being an artificial or purely theoretical philosophical problem. It accurately describes the fact that we experience ourselves both as self-conscious agents and as physically pre-structured entities situated both temporally and spatially (cf. Rinofner-Kreidl 2003b, pp. 125–205). A phenomenological critique of naturalism is based on vigorously acknowledging *the legitimate naturalization of consciousness*. This naturalization turns into a kind of *naturalism* which has to be refuted if one falsely ignores its »natural« limits (cf. Husserl 1952a, pp. 297, 346). The latter are brought to light by analysing the human mind as a complex phenomenon which comprises different ontological layers. The term »naturalization« rightly indicates the inseparability of nature and mind in every human consciousness. Strong naturalistic theories, on the contrary, annul the multi-layered structure of conscious-

ness in favour of describing its physiological organization as if the achievements of consciousness could be exhaustively grasped by confining oneself to its physiological foundation.

Thomas Metzinger's self-model theory takes notice of the physiological conditions which give rise to those occurrences we are accustomed to name »consciousness«. His *neurophenomenology* (Metzinger 1999b, p. 392) is meant to replace traditional phenomenology. It claims to give a naturalistic reformulation of subjectivity centring round phenomenal consciousness and cognitive self-reference. Metzinger's intention is not to eliminate subjectivity. Instead, he wants to take our first-person perspective seriously (cf. Metzinger 1999a, p. 9). Describing phenomena such as experiencing sensational contents, remembering, focusing one's attention on something or behaving like a self-conscious agent, Metzinger assumes that the representational contents of the mind are correlated with brain activities. Because this correlation is interpreted in terms of supervenience, the author believes his theory offers a non-reductive naturalization of the human mind. In particular, its naturalistic character is said to consist in treating consciousness in all its manifestations as a natural phenomenon which can be explained by natural means in reference to natural history (Metzinger 1994, pp. 42–43; Metzinger 1999a, pp. 18, 23–24, 117, 229). If we follow Metzinger in equating naturalistic theories with empirically founded theories, we are dealing with a weak and trivial conception of naturalism. Looking for more substantial statements, we learn that there are two which, according to Metzinger, are essential to naturalistic theories. Theories of this kind refuse both the idea of material *a priori* knowledge and the subject/object distinction as a basic category of our theorizing about the human mind. The former moment manifests itself in a bottom-up analysis that explains structural and procedural features of consciousness as resulting from self-organizing processes. These processes may be said to stand without need of any non-inherently functioning principles of regulation and control. When Metzinger argues in favour of his naturalistic theory, he proceeds from the alternative *naturalism vs. essentialism (sc. metaphysics) or metaphysics of subject vs. self-organizing natural systems*. He ignores the fact that a phenomenological approach cannot be subsumed under this alternative. Rather, it gets all its impulse and vigour from the attempt to open up a third way between naturalism and essentialism. This third way rests on the attempt to faithfully and painstakingly describe our lived experience.

Metzinger is not concerned with philosophically querying phenomenality in accordance with Husserl's *First Philosophy*. He does not take a phenomenological point of view according to which consciousness is considered the general medium of appearance. Consciousness, rather, is subject to a special methodical grip which replaces its intentional structure by internally produced hierarchies of informational data. These data are investigated with a view to their physical realization. Metzinger's project of presenting an empirically corroborated theory of the human mind is composed of a naturalization of *qualia* and a naturalization of intentionality. In his theory, »phenomenal consciousness« both refers to *Qualia* and to those operations which engender a *phenomenal ego*. As for the latter, it is particularly the sub-symbolic constitution of the phenomenal attribute of subject-centring which is of theoretical interest. *Qualia* are sensational contents whose quality seems to exclude any objective description, for the specific mode of being given to someone is constitutive of the quality at issue (cf. Metzinger 2000b). *Qualia* are considered simple, indivisible mental states possessing a

peculiar phenomenal content, namely »a subjective quality of experiencing which is accessible only to the person who has the experience in question« (Metzinger 1996a, p. 323). According to a phenomenological approach, »phenomenal consciousness« refers to two associated aspects of intentional experiences, namely: a) the appearing of x, and b) the present, living-through of the experience directed at x. Consequently, intentionality and phenomenality cannot be discussed separately. Contrary to this, Metzinger takes intentionality and phenomenality as representing two issues which can be treated independently from one another.

Metzinger holds that the representational character of contents is exclusively grounded in the causal roles which these contents fulfil as parts of an information processing system. Representation, therefore, is not an intrinsic character of contents of a special kind. Rather, representation has a functional role. The fundamental difference between Metzinger's concept of representation and the phenomenological concept of intentionality may be summarized as follows.

i) Information flows into the system from outside. Information processing, therefore, is referred to as »inverted intentionality« (Metzinger 1999a, pp. 128–129). The cognitive scientific concept of representation is based on the assumption that there are causal relations holding between an information processing system and its environment. According to a phenomenological concept of intentionality, intentional and real relations are irreducibly different insofar as the former may not be explained by referring to the latter. Of course, causal relations can concomitantly occur with intentional relations, for instance in the case of sensual perception. Nevertheless, intentional relations constitute a meaning content which cannot be reduced to physical or physiological issues, e.g. relations of stimuli and reaction.

ii) The phenomenological concept of intentionality combines two aspects, namely the directedness to an object, *via* some intentional content, and an immediate awareness of my actually living through the intentional experience in question. Metzinger eliminates the latter moment in favour of a conception that integrates phenomenal states into a self-model and a world-model by means of self-referentially structured meta-representations. Self-models occur whenever a part of the system analogically represents the system as a whole. Self-models do not involve any propositionally structured self-reference. Those parts of self-models which, in principle, can be represented as contents of phenomenal consciousness, *via* some meta-representation, are called »mental«. A phenomenal self is that part of the mental self-model which is the actual content of phenomenal consciousness (Metzinger 1999a, pp. 158–159). *Subjectivity* is a quality of complex information processing systems which manifests itself if and only if the system succeeds in embedding a self-model into its model of reality (Metzinger 1999a, p. 204).

The cognitive scientific conception of representation rests on a dual structure of internal data-processing and outer reality, one which is induced by Metzinger's concept of information. It implies a third-person perspective. The basic concepts of Metzinger's theory presuppose what this theory claims to show, namely that it is possible to naturalize our folk-psychological account of first-person perspective. A human first-person perspective precisely manifests itself in the intrinsic relatedness of consciousness and world, the latter being interpreted in terms of meaningful experiences. Due to its conceptual and methodical framework, Metzinger's self-model theory cuts off this intimate relation between consciousness and world. Metzinger

refutes Husserl's (and every other »classical«) conception of intentionality. In doing so, he ignores that a phenomenologist's and a cognitive scientist's idea of consciousness radically differ from each other. Thus, it is not surprising that he considers Husserl's phenomenological idealism an incomprehensible and obscure position. Among other things, Metzinger's critique is misguided with regard to the following issues.

- According to Metzinger, intentional experiences are present whenever an act (noesis) is directed to an intentional content (noema) as its object. In this view the mediating function of intentional contents gets lost. Whereas Husserl emphasizes that intentional contents are not the objects of our concern whenever we live through present intentional experiences (although we can turn our attention to them in subsequent acts of reflection), Metzinger implicitly takes this reflective turn to be constitutive of the »naïvely« realized intentional relation. From the point of view of phenomenology, this amounts to reifying intentional contents and consciousness respectively, since the fundamental reality of accomplishment (*Vollzug*), that is, of performing intentional experiences, is eliminated. Metzinger, however, denies any reifying interpretation.

»Content is not a mysterious type of *thing*, but an abstract property of a highly fluid and complex cognitive dynamics ...« (Metzinger 2000a, p. 302 (fn. 3)).

Representational contents must be described as

»... an aspect of an ongoing process and not as some kind of abstract object«. (Metzinger 2003b, p. 358)

- According to Husserl, intentional contents are both aspects of an ongoing process of experiencing *and* abstract objects, depending on whether they are considered as actually functioning or referred to by means of reflective acts (noematic reflection). Holding this view, requires, of course, that we note Husserl's purely formal understanding of object: »everything upon which we may predicate«. It furthermore requires us to acknowledge that living through actual experiences and reflecting on experiences are fundamentally different act qualities and that reflection presupposes actual experience. If we talk about a »*mediating* function« of intentional contents, this refers to the fact that every reference to an object involves some intentional content which determines the reference at issue. The relatedness of consciousness and world, *if analysed theoretically*, shows itself to be realized *by means of these contents*. This obviously does not involve a noematic *mediator* in accordance with the idea that consciousness and world are separate spheres of reality in need of being brought together by some ontologically mysterious *noema* entity representing an »outer reality«. This view certainly does not correspond to our experience of being intentionally related to something. In other words: it is ontologically and epistemologically harmless to refer to a mediating function of intentional contents, if this is interpreted in terms of the above mentioned one-sided genetic dependence of acts of reflection upon preceding accomplishments. *Acknowledging the mediating function of intentional content is equivalent to denying the representational character of our intentional relations.*
- Phenomenology does not replace real objects with intentional objects. Distinguishing intentional objects from real ones neither implies nor requires that we consider the former as fictitious. On the other hand, if intentional experiences are conceived of as functional states which represent aggregates of causal relations, we should say that intentional contents

represent presently given objects *because* mental states are determined by something different from them and located outside them (Metzinger 1999a, p. 17). However, in cases of remembering, imagining future events, fantasies, hallucinations and similar mental phenomena (»mental simulations«), it is obvious that there is no represented object as part of the present environment (cf. Metzinger 1999a, p. 66). Following Metzinger, therefore, we have to interpret intentional experiences on a large scale as fantasies which are not caused by any external occurrences.

- A phenomenological conception of intentionality does not assume a privileged first-person accessibility in terms of a full self-referential transparency of consciousness. Arguing like this is based on a dogmatic idea of givenness and a misguided conception of inner certainty (cf. Rinofner-Kreidl 2000, pp. 178–203; 387–412).

### III. A Scientific Model of Phenomenal Transparency: the Illusive Self

Phenomenal transparency (PT) is one of the basic features of the human mind. Transparency is a property of phenomenal representations in a sub-symbolic medium, i. e. of non-linguistic entities (Metzinger 2003b, p. 363). It is »a property of active mental representations satisfying the minimally sufficient constraints for conscious experience to occur« (Metzinger 2003b, p. 355), namely being presently activated and being integrated in a global model of the world.

»[Phenomenally] transparent representations are precisely those representations the existence of whose content we cannot doubt.« (Metzinger 2003c, p. 563)

PT is important with respect to explaining the achievement of cognitive self-reference that always is reference to the phenomenal content of a transparent self-model (Metzinger 2003b, p. 385). Why do transparent states emerge in information processing systems?

»(W)hat makes mental representations transparent is the *attentional unavailability of earlier processing stages in the brain* for introspection.« (Metzinger 2003b, p. 356)

PT is problematic with regard to its epistemological implications. PT is meant to elucidate a thesis which may be designated the »illusion thesis« (IT): The subjectivity of the mental results from a kind of self-confusion. A phenomenal ego appears because the system, in a certain respect, is deprived of information. This possibly is the most important insight of cognitive science concerning philosophical anthropology (Metzinger 1998, p. 361). We may expound this insight as follows:

(IT) The unity of our phenomenal self is a representational fiction (Metzinger 1996b, p. 152). The fact that we are faced with a fiction cannot be realized at the phenomenal, personal level of consciousness (Metzinger 2003b, p. 363; Metzinger 1994, pp. 50–51). Semantic transparency is responsible for our pre-reflexive self-acquaintance (Metzinger 1998, p. 360): A *phenomenal first-person perspective* emerges whenever a system is not able to recognize its self-model as a model. Only from the point of view of sub-personal information processing, it is possible to recognize the illusion and to explain why it necessarily occurs, given information processing systems of a certain complexity. Cognitive science helps us to understand why phenomenologists inevitably fall victim to the *myth of the given*. What phenomenologists naïvely grasp as immediately given truly is a construction based on neural processes.

The argument in favour of IT may be expounded in the following manner. What, in everyday experience, is given to us immediately, is interpreted in a naïvely realistic mode. Under normal conditions, we have the strong, albeit subjective, feeling that we are in direct contact with both ourselves as experiencing subjects and with the world. From a scientific point of view, it turns out that what presents itself as immediately given actually results from the inability of human consciousness to simultaneously grasp those rapidly running off brain processes that lie behind the experiences in question (cf. Metzinger 1999b, p. 401). As soon as we come to know this time lag inherent in all our thinking, perceiving, remembering or desiring, we cannot stick to the former naïvely realistic interpretation of ourselves and the world. As Metzinger argues, with a view to the Cotard syndrome, this implies that »if a human being's self-model became fully opaque, then this person would experience herself as non-existent« (Metzinger 2003f, p. 21). Abandoning naïve realism, we recognize that what actually happens when we live through theoretical and practical intentions is that some special system operations occur. Strictly speaking, there *is* no intention directed to an object »out there«. There is rather an informational process taking place. Intentional relations are substituted by an internal determinism of the information processing system. Metzinger's representationalism implies a constructivistic bias which cannot be challenged within the framework of his theory (cf. Metzinger 1996a, p. 622).

According to Metzinger, traditional theories of intentionality hold an incorrect idea of how the mental modelling of representational relations takes place. They erroneously attribute intentional experiences to a phenomenal self-model instead of attributing them to the brain that engenders the self-model in question. Referring to the latter, we do not find individual causal relations. A self-model exclusively grasps the final products of these underlying processes. From this Metzinger draws the conclusion that there is no consciousness which is directed to the world. There are, rather, complex physical occurrences determining mental models (Metzinger 1999a, pp. 128–129 (fn. 212)). This amounts to a naturalization of the cogito. Whereas Descartes took our thinking to be inseparable from our ego, we realize nowadays that, according to Metzinger, the ego is nothing but a thought which depends on a physical system *thinking this thought*. This system, for instance the brain of a biological organism, is the thinking thing (Metzinger 1999a, pp. 154–155).

From the above, it is clear that IT requires a highly problematic supplementary thesis which I call the »hidden agent thesis (HAT)«:

The true cognitive agent of our intentional experiences is not the phenomenal ego represented by a relatively stable and coherent self-model. The true agent is our brain.

Mental states do not represent their internal constructional genesis. They lack the quality of *Gewordenheit* (Metzinger 1996b, p. 143). The temporality of the underlying processes does not enter the representational content. Accordingly, IT may be reformulated as follows.

The illusive character of (at least some) representations is due to the fact that their neurological constitution is not part of the representational contents in question.

Let us call this T1. T1 involves a *fallacy of latency*. In order to explain why T1 is fallacious, we must now take up the problem of transparency as it presents itself in non-scientific contexts.

#### IV. A Phenomenological Model of Phenomenal Transparency: the Double Aspect of a Self's Behaviour (Accomplishment and Reflection)

Everyday experience is acquainted with a basic transparency of values, interests, purposes and so on which is indispensable for our thinking and acting. Whenever these issues are made, the objects of concern two things happen. First, they lose their immediate and reliable guiding function because we recognize that there are alternative values, interests, and purposes which could or should be equally realized. Secondly, the intentional objects of the experiences in question change as soon as we stop thinking and acting in a »naïve« mode and start reflecting on the rationales of our thinking and acting. Given that our hierarchies of values, interests, and purposes cannot be absolutely justified, the above reasoning leads to this: in order to think and act something, i.e. problematic ideas, must be kept in latency.

On the other hand, asking whether our thoughts and actions are *reasonable* involves a reflective turn. For the present purposes we may leave untouched what it means to give sufficient reasons for our acting and how we could hope to practically implement our principles of acting. With a view to the issue of transparency, it may suffice to note that in order to act smoothly these principles have to be made our own in a manner that allows for an »invisible« or unnoticed functioning. The fact that  $x$  functions latently does not indicate that  $x$  lacks any rational foundation. It might, just as well, turn out that it is perfectly rational to act  $x$ -like. In the absence of any reflection, practical (as well as theoretical) transparency leaves entirely open whether we are faced with an irrational behaviour or with a rational one *dropped into a special mode of passivity*, where passivity normally includes, as Husserl says, some hidden rationality. Acknowledging latent functions is neutral to epistemological concerns. The latter cannot be formulated as long as the intentional contents in question have not been made explicit. The occurrence of intentional experiences, as far as we know, depends on various latent functions of, for instance, biological organisms. However, describing latent functions does not mean giving a sufficient or even relevant description of the intentional experiences, e.g. thinking about something, whose occurrence is owing to the latent functions at issue. For instance, it is a reasonable hypothesis worth of being empirically tested that

»... [identity] disorders, while being diagnosed on the *personal* level of description, result from subpersonal disintegration« (Metzinger 2003f, p. 24).

Nevertheless, trying to explain the occurrence of pathological types of experience by referring to neurophysiological states or functional relations taken to be realized in these states, does not and cannot help us to gain a better, therapeutic understanding of what it means to have experiences of these kinds.

As we have seen above, relating to the mediating function of intentional content, there is also a transparency thesis implicit in a phenomenological conception of intentionality. When Husserl analyses the mode of givenness of other minds, he introduces the following analogy. The apperception of intentional contents that I attribute to other persons' mind is achieved *via* some bodily appearance in a way similar to the apperception of meaning which is achieved *via* some linguistic sign, »meaning« understood here in a

narrow logical sense (Husserl 1952a, p. 240). My body as well as the body of other persons functions transparently with regard to the apperception and communication of intentional contents. The phenomenological transparency thesis illustrates how phenomenologists steer clear of the alternative of a metaphysical essentialism and naturalism. Analysing the mind/body-relation, we have to recognize that, engaged in our life-world practice, we normally do not experience any separation of mind and body. According to Husserl, gaining sympathetic understanding (*einfühlendes Verstehen*) of other persons means to grasp their bodies with a view to inherent meaning structures manifesting a unity of sense. Every perception of a person is founded on an apperception of objective intentional content (*objektiver Geist*). On the other hand, there is no apperception of the body as bearer of the psychical in terms of a physical object which had to be supplemented by something different, as if it were taken

»... as something related or connected to something else. It, rather, is a higher-level objectification superimposing on a lower-level one to the effect that there results a unified object. Without referring to any kind of connection *which would imply separation*, this synthetic object comprises a lower- and a higher-level-constitution *being distinguished only subsequently*. The unity which is given in course of apperceiving some spiritual being can be distinguished into body and sense *by changing the apperceiving attitude.*« (Husserl 1952a, p. 244, emphasis mine, SR)

The so-called mind/body-problem arises if we theoretically reflect on the possibility of apperceiving other minds. However sophisticated this reflection may be, it always refers back and remains embedded in our experiential practice. *Acknowledging the transparency involved in this original practice amounts to overcoming the Cartesian starting-point of our theorizing both with regard to the original mind/body-unity and the original community of self and others.* Phenomenologically viewed, there is no need to naturalize the mind as long as our theoretical representations of the human mind do not go beyond its true nature, that is as it presents itself in primordial experience, in favour of introducing some dubious spiritual entity.

The phenomenological transparency thesis differs in an essential way from Metzinger's transparency thesis. First, it is not connected with any illusion thesis. If we explicitly refer to the functioning of bodily appearance and behaviour or linguistic signs, an act which we may accomplish at any time by turning our attention, this does not alter the epistemic appraisal of the intentional content at issue. Acknowledging the indispensability of latent functions does not alter the phenomenal character of those previous experiences that have been effected because of the latent functions in question. Second, and in connection with the epistemic neutrality of reflecting on transparent moments, the phenomenological transparency thesis does not annihilate the distinction between living through experiences and reflecting on experiences. On the contrary, phenomenological transparency implies that this distinction essentially belongs to human consciousness. Third, whether something is explicitly grasped or given transparently, whether it is, to use Husserl's terms, given in a manifest (*patent*) or latent manner, depends on the thematic attitude of the experiencing subject. In order to understand transparency, we must inquire into the attitude object correlations that belong to the intentional life of persons, instead of analysing special attitudes, e.g. attention, *in terms of information processing*. In order to distinguish latent and manifest functions, we must make reference to changing attitudes or perspectives, modes of querying. Nothing is *in itself* latent or manifest. The appearance of whatever thing and moment depends on

particular circumstances and conditions on the side of the subject, as well as on the side of the object which must be specified. Every single attitude brings to light particular objects and processes and, simultaneously, makes disappear other objects and processes. For instance, if we decide to neurologically inquire into the genesis of phenomenal transparency, we unavoidably lose out of sight the intentional life of persons. IT forfeits its *prima facie*-plausibility if we realize that every scientific finding rests on some particular attitude which fulfils a visualizing and de-visualizing function in the above sense. This function is vital for our scientific division of labour. Nevertheless, we are led astray if we forget its operational implications.

#### V. Scientifically Shaped Experience *versus* Primordial Experience: the Representationalist Predicament

The latency of neural processes implied in Metzinger's PT cannot be conceived of as a structure of experience. There is no subject for whom the processes in question could be made explicit *as structuring her experience*. The brain (or: the limbic system) does not constitute a self in any relevant theoretical or practical sense. It cannot constitute a self because this requires meaningful, i.e. semantically interpreted experiences that involve reference to (hierarchies of) wants, desires, values, interests, and so on. The mere running off of neural processes does not constitute experience. The processes occurring in my brain are not part of my experience (cf. Husserl 1952a, pp. 164, 218, 230–231; Ricoeur 1996, p. 164). IT and HAT call for self-application. Every cognitive scientist who investigates the human mind, and thereby performs particular intentional experiences, too, must be subject to IT and HAT. She has to consider herself as a brain that interprets informational data realized in some mind whose »true agent« is a brain. From this point of view it may become common to talk about brain-to-brain communications in cognitive science. Arguing like this, however, involves a categorical mistake. Neural processes do not think, communicate, regret their moral imperfection or enjoy some magical erotic moments.

If there are electrodes fixed to my head which transmit the occurrences in my brain to a recorder that simultaneously visualises them, it does not follow that I am, in any strict or interesting sense, able to experience my own brain. Methodically disciplined modes of experience which are operative in variable experimental designs must not be confused with our ordinary experience (cf. Heidegger 1996, pp. 188, 196–197). Metzinger naturalizes primordial experience by supposing that *first-person perspective and third-person perspective are equivalent attitudes which can be changed arbitrarily according to different objectives* (Rinofner-Kreidl 2004). Consequently, we are free to describe our mind from either an external point of view, i. e. as a succession of neurological states, or an internal point of view, i. e. as the experiences we go through (Metzinger 1996a, p. 256). Here again it is obvious that the naturalization of subjectivity which is expected to be the outcome of Metzinger's theory has been smuggled in from the outset (see section III). It is implied in the basic concepts of his self-model and its mode of questioning empirical data. Contrary to this, a phenomenologist insists that any talk about »my brain« requires that I have an experience of my own body involving a first-person perspective. Without this original bodily

experience, I could never conceive of a succession of brain states as belonging to me. Taking first-person perspective and third-person perspective on a par with one another, tacitly assumes that brains can be considered self-sufficient entities whose contingent realization in human bodies, including modes of behaviour which cannot be explained »bottom-up«, can be neglected with a view to special theoretical purposes.

A functionalistic view, committed to the thesis of multi-realizability, cannot do justice to our subjective experience of embodied mind. What it means to leave aside the intrinsic »worldliness« of human consciousness can be explained in relation to a diverse set of issues. Among these is the peculiar weakening of the mind/body unity owing to a functionalistic view, i.e. the displacement of our natural experience of embodiment. In a similar way, the loss of world in Metzinger's theorizing manifests itself as a total neglect of intersubjectivity with a view to the constitution of self-reference and, consequently, self-referentially based self-consciousness. Furthermore, Metzinger disregards our ability to grasp immediately the situational horizon of whatever single experience we encounter. It has been rightly argued that scientific models that take consciousness to consist in information processing inevitably fail to consider the »enworlded« nature of consciousness. This is due to the fact that it is impossible to transform our natural ability of synthetically grasping more or less complex situations into some objective system of data by means of which one could exhaustively specify the concepts and rules to be applied in concrete cases (Dreyfus 1985, pp. 214–218). The belief that this can be done, applied to artificial intelligence, could be called »computer-Cartesianism« (cf. Dreyfus 1985, p. 218). Equally, we may talk about a »brain-Cartesianism« implicit in those cognitive scientific theories which, due to their representationalism, adhere to a rigid dualism of outer reality and internal system of informational data (see the discussion about our concept of the world in section VII). Both the external and the internal eludes our natural experience.

From another point of view, the sterility of Metzinger's theory concerning the demands of intersubjectivity and situational embedding is due to the complete lack of a practical dimension. Metzinger's naturalistic theory of self-modeling exclusively explains the emergence of phenomenal self-consciousness without taking into account the social dimension of subjectivity (Metzinger 1999a, pp. 175, 225). This dimension is said to be negligible insofar as the physical structures of data are deprived of any semantic content (Metzinger 1999a, p. 251 (fn. 10)). Notwithstanding this severe restriction of his field of investigation, the author does not reckon with any loss of sense with regard to his naturalized subjectivity. Generally, and quite amazingly, it is expected that future research in the field of cognitive science will succeed in incorporating intersubjectivity. And yet, how could one expect to adequately describe or explain subjectivity from a solipsistic point of view and, later on, expand one's concern to a social dimension considered as a supplementary field of experience which is not, from the outset, involved in the constitution of subjectivity?

Metzinger's attempt to naturalize subjectivity is based on a reification of consciousness. This reification results from eliminating the distinction of actually living through experiences and being reflectively directed at one's experiences. The representationalism of Metzinger's self-model theory does not leave any room for referring to immediately present experiences. For methodical reasons, Metzinger holds that what cannot be represented as

conscious content on the neurophenomenological level of description is not a relevant part of human experience. This assumption occasionally turns up when Metzinger discusses the methodical implications involved in the question of how philosophers and psychiatrists can cooperate in analysing phenomena of identity disorder.

»'Personal identity' can be either a complex theoretical concept or a *concrete [form] of subjective experience, a conscious content* – but it never is a thing, neither in the brain nor anywhere else. To put the point differently: Psychiatrists must stop being naïve realists about personal identity.« (Metzinger 2003f, p. 3, emphasis mine)

Here we have the tacit presupposition of Metzinger's self-model: There are no »formal« aspects of human experience which cannot be (neurophenomenologically) represented as conscious contents. Depending on whether or not we agree to this general thesis, we achieve completely different ideas of subjectivity. It is characteristic of a phenomenological approach to deny the above thesis (Rinofner-Kreidl 2003b, pp. 215–221). The peculiar character of subjectivity ultimately lies in the fact that there are »formal« aspects of our experience which resist representational objectification on whatever level of description. Metzinger holds that this is true with regard to the phenomenal level of description, as it is interpreted in folk psychology, but false with regard to the neurophenomenological level of description. Consequently, a pivotal question of the naturalism debate is how different levels of description are related to each other. However, talking about »levels« of description is misleading. It wrongly suggests that what neurophenomenologists and phenomenologists do is describe the same phenomena from different points of view.

In relation to the superimposing structure of meta-representations, which enables the system to re-interpret and overcome phenomenally transparent states, we may summarize Metzinger's account as follows. Experiences are part of our cognitive life only if they have been transformed into or acquired in terms of meta-representational contents. Contrary to this, a phenomenological approach emphasizes *the practical moment of accomplishment* (living through experiences). From a phenomenological point of view, it is evident that Metzinger de-temporalizes our primordial self-consciousness which is said to consist in a hierarchical structuring of contents by means of meta-representations. It is interesting that Metzinger, *vice versa*, accuses phenomenologists of eliminating the dynamic character of consciousness (»phenomenological fallacy«: Metzinger 1999a, p. 161) by disregarding the neurological level of description. This objection results from two assumptions which must be dismissed. First, the succession of occurrences in the brain is equated with the original temporality of consciousness based on the subjective experience of before/after relations, which are not yet interpreted in terms of objective time units. Second, the mediating function of intentional contents is ignored (see above section III).

## VI. Meta-theoretical Issues Involved in the Naturalism Debate

Metzinger's naturalization project revolves round the attempt to conceive of subjectivity in terms of representational relations. Contrary to this, a phenomenologist argues that representation cannot be considered the most fundamental, original mode of consciousness. Given this essential dis-

agreement concerning the idea of consciousness, the most basic question seems to be whether cognitive scientists and phenomenologists actually refer to the same phenomena or not. Do they explain the same phenomena in incompatible ways or do they talk about different things? We obviously cannot soothe this issue by arguing that the opponents talk about different aspects of reality so that, in a certain sense, we might be satisfied with both approaches. This does not do because our idea of reality is part of the issue. If we follow Metzinger in conceiving of intentionality in terms of causal roles, it is unclear how we could ever reach the idea that the world is *not* the sum total of all things, and instead an encompassing meaning-horizon without which we were unable to encounter things. It does not suffice to consider the holistic character of reality (Metzinger 1995), if this holism, again, is interpreted according to the functionalistic approach of Metzinger's »bottom-up« neurophenomenology. If we proceed from the scientifically reduced, non-natural idea of the world which manifests itself in Metzinger's self-model theory, it may well be that we can describe how our brains are related to a virtual or real world. Nevertheless, this does not say anything about our natural experience of *the* world. It only shows that in constructing theoretical models we are free to talk about a world or certain aspects of reality. This is owing to the fact that the definition of »world« or »reality« depends on internal criteria of the models in question. From this point of view, the main concern of the naturalism debate is how we should consider theoretical models to be related to our life-world practice. Can we really leave it to some special scientific model to determine what »world« or »reality« means? Or, must we rely on our primordial experience which always operates on condition that we already have an understanding of world and reality? Should we not acknowledge this fundamental status of primordial experience, how, then, could we claim to critically assess *different* models of reality resulting from different theoretical approaches? Genetically viewed, our natural understanding of the world is already involved if we are told that to »have an ontology is to interpret a world« and that »the brain, viewed as a representational system aimed at interpreting our world, possesses an ontology, too.« (Metzinger 2003c, p. 549) In methodical terms, this amounts to a

»... mathematical model describing the phenomenal ontology of the human brain – i.e. that which *exists* according to conscious experience – in a precise and empirically plausible manner.« (Metzinger 1995, p. 427).

If a phenomenologist advances the view, as I do, that representationalistic theories distort or even eliminate those phenomena they claim to explain, what then, should we think about the naturalism debate insofar as theories of this kind are concerned? We certainly cannot expect that: a) scientific methods are neutral with regard to the objects investigated, and that b) the naturalism debate can be decided by comparing the efficiency of diverse (scientific) methods of analysing and explaining phenomena. Every statement concerning methodical efficiency tacitly implies some conception of how our theorizing is related to the world as it presents itself in everyday life. Applying methods always involves a certain structuring of its correlating objects. This being the case, it is clear that the dispute between naturalists and anti-naturalists cannot be settled at the level of describing and explaining phenomena. To describe or explain phenomena and to discuss the methodical and conceptual framework of our descriptions and explanations are quite different activities which should be clearly distinguished. How-

ever, the results of the former activity might turn out to be problematic due to ignoring the difficulties brought to light by the latter. Both a) and b) suggest that if we phenomenologically challenge projects of naturalizing the human mind we have to focus on the issue of how theory and practice are interrelated. Thus, asking on what *theoretical* grounds it might be possible to adequately discuss naturalizations of subjectivity is still insufficient if it does not comprise a radical reflection on our idea of theory.

### References:

- Dreyfus, H. L. (1985), *Grenzen künstlicher Intelligenz. Was Computer nicht können*. Königstein/Ts.: Athenäum.
- Heidegger, M. (1996), *Einleitung in die Philosophie*. Frankfurt: Vittorio Klostermann (= Collected Works Vol. 27).
- Heidegger, M. (<sup>18</sup>2001), *Sein und Zeit* Tübingen: Max Niemeyer.
- Husserl, E. (1952a), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Zweites Buch: Phänomenologische Untersuchungen zur Konstitution*. Ed. by M. Biemel. Den Haag: Martinus Nijhoff (= Husserliana Vol. IV).
- Husserl, E. (1952b), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Drittes Buch: Die Phänomenologie und die Fundamente der Wissenschaften*. Ed. by M. Biemel. Haag: Martinus Nijhoff (= Husserliana Vol. V).
- Husserl, E. (1962), *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie*. Ed. by W. Biemel. Den Haag: Martinus Nijhoff (= Husserliana Vol. VI).
- Husserl, E. (1985), *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*. Ed. by L. Landgrebe. Hamburg: Felix Meiner.
- Husserl, E. (2001), *Natur und Geist. Vorlesungen Sommersemester 1927*. Ed. by M. Weiler. Dordrecht/Boston/London: Kluwer Academic Publishers (= Husserliana Vol. XXXII).
- Metzinger, T. (1994), »Schimpansen, Spiegelbilder, Selbstmodelle und Subjekte«, Krämer, S. (ed.), *Geist – Gehirn – künstliche Intelligenz. Zeitgenössische Modelle des Denkens*. Ringvorlesung an der Freien Universität Berlin. Berlin, New York: De Gruyter, pp. 41–70.
- Metzinger, T. (1995), »Faster than Thought. Holism, Homogeneity and Temporal Coding«. In: *Conscious Experience*. Thoverton: Imprint Academic/Paderborn: mentis, pp. 425–461.
- Metzinger, T. (<sup>3</sup>1996a) (ed.), *Bewußtsein. Beiträge aus der Gegenwartsphilosophie*. Paderborn et al.: Ferdinand Schöningh.
- Metzinger, T. (1996b), »Niemand sein. Kann man eine naturalistische Perspektive auf die Subjektivität des Mentalen einnehmen?«, Krämer, S. (ed.), *Bewußtsein. Philosophische Beiträge*. Frankfurt: Suhrkamp, pp. 130–154.
- Metzinger, T. (1996c), »Ganzheit, Homogenität und Zeitkodierung«, Metzinger, T. (1996a), pp. 595–633.
- Metzinger, T. (1998), »Anthropologie und Kognitionswissenschaften«, Gold, P., Engel, A. K. (eds.), *Der Mensch in der Perspektive der Kognitionswissenschaften*. Frankfurt: Suhrkamp, pp. 326–372.
- Metzinger T. (<sup>2</sup>1999a), *Subjekt und Selbstmodell. Die Perspektivität phänomenalen Bewußtseins vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation*. Paderborn: Mentis.

- Metzinger, T. (1999b), »Präsentationaler Gehalt«, Esken, F., Heckmann, D. (eds.), *Bewußtsein und Repräsentation*. Paderborn: Mentis, pp. 377–405.
- Metzinger, T. (2000a), »The Subjectivity of Subjective Experience: A Representational Analysis of the First-Person Perspective«, T. Metzinger (ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: The MIT Press, pp. 285–306 (<http://www.philosophie.uni-mainz.de/metzinger/publikationen>, 20. 6. 2004).
- Metzinger, T., Walde, B. (2000b), »Commentary on Jakob's *Ineffability of Qualia*«, *Consciousness and Cognition* 7, pp. 353–362.
- Metzinger, T. (2003b), »Phenomenal transparency and cognitive self-reference«, *Phenomenology and Cognitive Science* 2, pp. 353–393.
- Metzinger, T., Gallese, V. (2003c), »The emergence of a shared action ontology: Building blocks for a theory«, *Consciousness and Cognition* 12, pp. 549–571.
- Metzinger, T., Gallese, V. (2003d), »Motor ontology: the representational reality of goals, actions and selves«, *Philosophical Psychology* 16, 3, pp. 365–388.
- Metzinger, T. (2003e), »The pre-scientific concept of a 'soul': A neurophenomenological hypothesis about its origin«, F.-M. Peschl, H.-D. Klein (eds.), *Haben Computer eine Seele? Der Begriff der Seele* Bd. IV. Würzburg: Königshausen & Neumann (forthcoming), pp. 1–36 (<http://www.philosophie.uni-mainz.de/metzinger/publikationen>, 20. 6. 2004).
- Metzinger, T. (2003f), »Why are identity-disorders interesting for philosophers?«, T. Schramme, J. Thome (eds.), *Philosophy and Psychiatry*. Berlin: de Gruyter (forthcoming), pp. 1–29 (<http://www.philosophie.uni-mainz.de/metzinger/publikationen>).
- Nagel, T. (1997), »Why We Can't Understand Thought from the Outside«, Nagel, T., *The Last Word*. New York, Oxford: Oxford University Press.
- Petitot, J./Varela, F. J./Pachoud, B./Roy, J.-M. (1999) (eds.), *Naturalizing Phenomenology. Issues in Contemporary Phenomenology and Cognitive Science*. Stanford/California: Stanford University Press.
- Ricœur, P. (1996), *Das Selbst als ein Anderer*. München: Wilhelm Fink.
- Rinofner-Kreidl, S. (2000), *Edmund Husserl. Zeitlichkeit und Intentionalität*. Freiburg, München: Karl Alber.
- Rinofner-Kreidl, S. (2003a), »Do Cognitive Scientists succeed in Naturalizing Free Will?«, Kanzian, C., Quitterer, J., Runggaldier, E. (eds.), *Persons. An Interdisciplinary Approach/Personen. Ein interdisziplinärer Dialog*. Wien: Österreichischer Bundesverlag, pp. 222–231.
- Rinofner-Kreidl, S. (2003b), *Mediane Phänomenologie. Subjektivität im Spannungsfeld von Naturalität und Kulturalität*. Würzburg: Königshausen & Neumann.
- Rinofner-Kreidl, S. (2004), »Das ‚Gehirn-Selbst‘. Ist die Erste-Person-Perspektive naturalisierbar?«, *Phänomenologische Forschungen* (forthcoming).
- Thompson, E., Lutz, A., Cosmelli, D. (2004), »Neurophenomenology: An Introduction for Neurophilosophers«, Brook, A., Akins, K. (eds.), *Cognition and the Brain: the Philosophy and Neuroscience Movement*. New York and Cambridge: Cambridge University Press (forthcoming).
- Varela, F. J./Thompson, E./Rosch, E. (1993), *The Embodied Mind. Cognitive Science and Human Experience*. Cambridge, London: The MIT Press.

Sonja Rinofner-Kreidl

## **Die Grenzen des Repräsentationalismus**

### **Eine phänomenologische Kritik von Thomas Metzingers Selbstmodell-Theorie**

*Die Theorie des Selbstmodells von Thomas Metzinger bietet einen Rahmen für die Naturalisierung subjektiver Erfahrungen, d.h. der Ich-Perspektive. Diese Phänomene lassen sich erklären unter Bezug auf Repräsentierungsinhalte, die, so heißt es, auf verschiedenen Bewusstseinssebenen interrelational und mit Gehirnaktivitäten korrelierend sind. Der Artikel beginnt mit der Ergründung von Naturalismus und Anti-Naturalismus als Skizze für den Hintergrund der Metzingerschen Behauptung, seine Theorie befände philosophische Spekulierungen über den Geist für nicht notwendig (I). Insbesondere wird Husserls phänomenologischer Bewusstseinsbegriff als unkritisch und unangebracht verworfen. Es wird sich herausstellen, dass diese Kritik falsch gerichtet war (II). Die Hauptmängel der Metzingerschen Theorie werden durch Referieren auf den Begriff der phänomenalen Transparenz beleuchtet (III), die mit der phänomenologischen Idee der Transparenz verglichen werden soll. Dann wollen wir unseren kritischen Horizont erweitern, indem wir einige Implikationen des Repräsentationalismus fokussieren werden, einschließlich der Reifizierung des Bewusstseins, des Gehirn-Cartesianismus und des Ausschlusses der sozialen Dimension (V). Schliesslich wollen wir unsere metatheoretischen Reflexionen auf die Naturalismusdebatte übertragen (VI).*

Sonja Rinofner-Kreidl

## **Les limites du représentationnalisme**

### **Critique phénoménologique de la théorie du moi modèle de Thomas Metzinger**

*La théorie du moi modèle de Thomas Metzinger offre comme cadre de naturalisation d'expériences subjectives le point de vue de la première personne. Ces phénomènes sont expliqués par référence aux contenus représentationnels, considérés comme corrélationnels à différents niveaux de la conscience et corrélatifs aux activités cérébrales. L'article commence par des observations au sujet du naturalisme et de l'anti-naturalisme en vue d'esquisser à grands traits les fondements des assertions de Metzinger selon lesquelles sa théorie rend superflues les spéculations philosophiques sur l'intellect (I). En particulier, le rejet de la conception phénoménologique de la conscience de Husserl comme inappropriée est peu fondée. L'article démontre que cette critique va dans est peu judicieuse (II). Les principaux défauts de la théorie de Metzinger y sont mis au clair par référence à la conception de transparence phénoménale (III), qui est comparée à l'idée phénoménologique de transparence (IV). C'est alors que nous y élargissons notre horizon critique en prêtant une attention particulière à certaines implications du représentationnalisme, y compris la réification de la conscience, le cartésianisme cérébral et l'exclusion de la dimension sociale (V). Finalement, nous appliquons nos réflexions métathéoriques sur la discussion au sujet du naturalisme (VI).*