

Abschlussarbeit

zur Erlangung des Magister Artium im Fachbereich 08
der Johann Wolfgang Goethe Universität
Institut für Philosophie

Thema

Mental content. Consequences of the embodied mind paradigm

1. Gutachter: Dr. Jasper Liptow
2. Gutachter: Prof. Dr. Martin Seel

vorgelegt von: Michael Sienhold

aus: Eisenach

Einreichungsdatum: 22. Februar 2010

Table of contents

Introduction.....	4
1 Crucial thoughts of the embodied mind paradigm.....	6
1.1 Cognitivist objectivist semantics versus embodied cognition	6
1.2 Lakoff/Johnson's notions of image schemata and basic-level concepts.....	8
1.3 Neural exploitation – how concrete and abstract concepts make use of the sensorimotor system (Gallese & Lakoff 2005).....	10
1.4 Sensorimotor coupling and a weakened version of the dual visual systems hypothesis	14
1.5 Body schema and body image – double dissociated phenomena.....	14
1.6 At a subpersonal functional (non-neural) level – the inverse model and the forward model	16
1.7 Embodied and embedded mind without extended functionalism.....	18
2 Conceptual tools: Representation, simulation, and presentation.....	18
2.1 A mind-dependent reality – conscious experience as a virtual world.....	19
2.2 Mental and phenomenal representation.....	22
2.3 Mental and phenomenal simulation.....	24
2.4 Mental and phenomenal presentation.....	25
2.5 Phenomenal mental models.....	27
2.6 Thoughts on the vehicle-content distinction.....	27
3 The mechanistic framework	31
3.1 The metaphysical mind-body-environment problem	31
3.2 Central properties of mechanistic explanations.....	31
3.3 Thoughts on multiple realisability.....	37
3.4 Some objections to other classical arguments against identity theories.....	44
4 Barsalou's theory of perceptual symbols systems (PSS).....	48
4.1 The symbol grounding problem and the Chinese Room	48
4.2 Feature maps and convergence zones (Damasio's hypothesis).....	49
4.3 Reconstructing PSS.....	51
4.4 PSS and dynamical systems theory: context sensitivity, embodiment, and temporality	59
4.5 A neurally embodied theory of language.....	60

4.6 Further empirical evidence for concept empiricism.....	65
4.7 Identifying PSS as a mechanistic explanation.....	71
4.8 Problems for the postulation of amodal symbols.....	72
4.9 Discussing objections to concept empiricism.....	73
5 Against computationalism (both symbolic and connectionist).....	81
5.1 A critique of symbolism.....	81
5.2 A critique of connectionism.....	84
5.3 Spivey's understanding of dynamical systems theory.....	86
6 Implications of embodied cognition in general and PSS in particular for classical philosophical questions concerning mental content	91
6.1 An embodied theory of intentionality	91
6.2 Criticizing some classical theories of intentionality.....	94
6.3 Narrow mental content	98
6.4 Nonconceptual mental content.....	100
6.5 Phenomenal perception is cognitively penetrable	104
6.6 Indeterminateness of conceptual processing.....	105
6.7 Criticizing the explanations a possible world semantics provides for non-synonymous co-extensional expressions and intensional contexts of beliefs	107
6.8 Eliminating or reformulating classical semantic notions: truth, reference, compositionality, inferential relations, and normativity.....	109
6.9 Methodological consequences of embodied cognition.....	114
7 Summary.....	117
References.....	123
Appendix.....	140

Introduction

The following work is premised on a view of philosophy that is not quite widespread. According to this the majority of philosophical questions has to be considered against the backdrop of empirical knowledge. Concretely, if philosophy consists in the explication of conceptual relations, then I think that those concepts have to be specified in terms of the respective empirical theories and on the basis of this one can explicate the relevant conceptual relations. The philosophical problems that are supposed to be solved are those that came up in the course of the history of philosophy. To give an example concerning the topic of the following work, the question whether nonconceptual content exists or not is determined subject to neuroimaging studies. Thus I proceed not only on the assumption that philosophy has to be compatible with empirical theories, but I also think that philosophy is said to make use of empirical information in order to elaborate theoretical positions.¹

The following considerations oppose the view that language is constitutive of all kinds of mental² content. I will argue for a non-linguistic understanding of concepts and for the ontogenetic grounding of higher-level cognition in phenomenal structures. Moreover, I will go on at a realistic and intersubjectively objective view of the mental.

The overall goal is to develop an exhaustive conceptualization of mental content in terms of embodied mental systems. One of the main principles of cognition I argue for is concept empiricism: All concepts are embodied in sensorimotor content. The vehicles of thought processing this content are perception and action mechanisms of the brain. (e.g., Varela et al. 1991)

The first Chapter aims at introducing crucial elements of embodied cognition. At first I will present Lakoff/Johnson's analyse of meaning and understanding in terms of image-schemata and basic-level concepts. In this context I will show how concrete and abstract concepts make use of the sensorimotor systems (what is called 'neural exploitation'). Three different notions of embodied simulation respectively residing in a certain functional cluster of the

¹ I think that the objection which is frequently made to neurobiologically grounded philosophical theories (that is, neurophilosophy), namely that the interpretation of neural processes presupposes a phenomenological and social prior understanding, is not conclusive. Why? Because even if those theories rely on a prior understanding, this does not imply that the prior understanding being used cannot be constituted by neurobiological processes. To this extent neurophilosophy may be considered as contributing to our self-awareness.

² In order to avoid misunderstandings resulting from terminological ambiguities, I define 'mental', 'mind', 'cognitive' and 'internal' as follows: I use 'mental' and 'mind' so that they are interchangeable. They name those episodes that can become globally available. We shall see later what this means. 'Cognitive' is only used if we name mental components whose activation is fully decoupleable from ongoing online cognition. That is, they have to be storable in long-term memory. 'Internal' processes stand for cognitive, mental and those processes that cannot become globally available.

brain are given. Subsequently, Milner/Goodale's two visual pathways hypothesis and the pervasive structure of sensorimotor coupling are brought in line. Then, the body schema and the body image are introduced as two functionally distinct components of the human brain. Finally the inverse and forward model are presented with the aim of giving a more or less non-neural functional determination of constituents of online processing.

In Chapter two a functional characterization of phenomenal mental processes is given: Consciousness is analysed in terms of availability for guided attention, behavioural control, and cognitive processing (global availability). The three concepts introduced – presentation, representation, and simulation – are *inter alia* characterized by the degree to which they are globally available. Beyond that, I will delineate each concept in non-phenomenal terms. The majority of mental processes takes place non-phenomenally.

The topic of Chapter three is a model of neuroscientific explanation ('mechanistic explanations') that reconsiders reduction, causation, emergence, identity, realisation, and multiple realisation. I think its highest merit is to emphasize the importance of making relations between different levels (e.g., functionally individuated brain regions and phenomenal experiences or observable behaviour) intelligible.

Chapter four deals with Barsalou's theory of perceptual symbol systems (PSS). The pivotal idea is to ground higher-level cognition in re-enactments of sensorimotor areas of the brain. After a detailed reconstruction of PSS, I will outline some behavioural and neurophenomenological evidence for embodied cognition in general and PSS in particular. A neurally embodied theory of language is also presented in this chapter. Finally, I will discuss objections to the entire project of concept empiricism.

The goal of Chapter five is to reveal the weaknesses of symbolism and connectionism as two versions of computationalism. These negative contributions are contrasted with a positive account of dynamical systems theory. It defines the mind in terms of an ongoing, real-time sensation-cognition-action process.

The aim of Chapter six is to spell out the implications of embodied cognition in general and PSS in particular for classical philosophical questions concerning mental content. At first I will conceptualize an embodied theory of intentionality. Linked to that I will criticize some classical models of intentionality, pre-eminently the one of Putnam/Kripke and the model of informational semantics. Subsequently, I will provide four positive contributions: justifying why PSS or embodied simulation fixes mental content narrowly; giving a version of nonconceptual content; explicating that perception is cognitively penetrable; and showing how PSS as a version of concept empiricism can account for the indeterminateness of

conceptual processing. Then I will criticize the explanations a possible world semantics provides for non-synonymous co-extensional expressions and intensional contexts of beliefs. The second last section aims to eliminate, or reconsider, truth, reference, compositionality, and normativity as classical semantic notions, as already done with amodal symbols and the computational version of propositions. Finally, I will elucidate the methodological consequences of embodied cognition – that is, argue for the unreliability of philosophical intuitions and formulate an epistemically relevant concept of possibility.

1 Crucial thoughts of the embodied mind paradigm

The current mind-body problem [...] is not *whether* minds are part of the natural world, but *how* they are. (Polger 2004, p. 1)

My intention is to outline pivotal assumptions made by the embodied cognition framework. It is important to note that this chapter is not mainly argumentatively-oriented, but rather aims at making the approach of embodied cognition intelligible. As a matter of course, I introduce distinctions which are used later in more argumentative contexts.

1.1 Cognitivist objectivist semantics versus embodied cognition

To make one's own profile clear it often seems appropriate to outline how it differs from the opposite position. Hence, I am going to briefly depict the crucial tenets of cognitivist objectivist semantics in order to subsequently present the basic philosophical ideas of embodied cognition.

(i) The meaning of mental states is based on a relation between symbolic, language-like mental representations and objective, mind-independent reality. These symbolic representations receive their content exclusively via their correspondence to external individuals, properties, and relations that are constituents of an objective world. (ii) Concepts are symbols that bear relations to other concepts within conceptual systems and correspond with individuals and categories of real and possible worlds. Concepts have to be quite general, since they represent what is common to various particular objects. The meaning of a concept is a function of the individuals to which it is applicable in all possible worlds. The meaning of a sentence-like proposition is a function of its truth-value in all possible worlds. (iii) Concepts are not bound to concrete embodied experiences of subjective minds. Their general, communicable, objective character requires that they are independent of any

concrete embodiment. Conceptual relations as objective relations are independent of the way they are intelligible within cognitive systems. What has been said so far implies that there is an objectivist view from which the correspondence between quasi-linguistic concepts and external states of affairs can be assessed. (iv) A theory of mental content must be able in principle to determine the truth conditions or satisfaction conditions of quasi-linguistic mental representations. The recursive character of a theory of mental content allows it to build larger true representations from smaller ones by means of logical connection. (cf. Lakoff 1987, pp. 157-84)

What is the opposite of this? On the basis of an experientialist account of cognition, meaningfulness is analysed in terms of preconceptual sensorimotor structures. Understanding is analysed in terms of meaningfulness. Truth in a psychological sense is analysed in terms of understanding the correspondence between cognitive mental representations and phenomenal experiences. Objectivity is analysed in terms of elucidating how we understand.

The external referents or objective states of affairs are only intelligible against the background of our conceptual system. To act as if the description of the truth or satisfaction conditions of language-like mental representations were objective and independent of our conceptual system – whose very content is in need of explanation – is highly questionable. As we will see later, one could even go as far as to say that such an objectivist conceptualization of mental content is due to a naïve realism that is suggested by our transparent phenomenal mental world model, which can be characterized at best in teleofunctionalist terms. I am drawn to the conclusion that cognitivist objectivist semantics considers nonconceptual visual experiences as being in contact with the external world. How could it otherwise be considered possible to provide quasi-linguistic tokens with meaning by relating them with objective descriptions – that rely on our conceptual system – of truth-conditions? Would one say that one can articulate truth or satisfaction conditions of language-like representations, if one appreciated the non-epistemic phenomenal character of visual perception? Would one speak of describing truth or satisfaction conditions, if one thought that the described referents or states of affairs were not phenomenally accessible? I don't think so.

What does it mean to say that meaningfulness is analysed in terms of preconceptual sensorimotor structures?

1.2 Lakoff/Johnson's notions of image schemata and basic-level concepts

The project undertaken by Lakoff (1980, 1987) and Johnson (1987) (1999) is presented here. In order to outline the sense of embodied understanding and meaning that Johnson and Lakoff have in mind, it is essential to elucidate the roles played by *kinesthetic image schemata* (or *embodied schemata*) and *basic-level concepts*. The experientialist strategy characterizes meaning in terms of experiences undergone by biological organisms. More concretely, the conceptual structures arise from bodily experiences themselves that are pre-conceptually structured. Our pre-conceptual experiences are structured at least two-fold: Through basic-level structures and kinesthetic image-schematic structures. Basic-level concepts directly correspond to preconceptual structures that are based on the part-whole structure in gestalt perception (e.g., the shape of a tiger quite rich in structure), our capacity for motor movement and mental imagery. These basic-level concepts should not be regarded as primitive, non-analysable, and unstructured building blocks of conceptual systems. Importantly, being structured and being non-basal do not coincide with one another. They are actually intermediate in conceptual systems – for example, chairs are subordinate to furniture and superordinate to office chairs. Yet what is critical is that they are human artifacts which are constructed so that our bodies can optimally interact with them. Human beings possess basic-level concepts for objects (e.g., tables, tigers, water (they correspond to the aspect of 'natural kinds' that is phenomenally visible with the naked eye) etc.), actions (e.g., eating, running, walking, etc.), and properties (e.g., small, cold, soft, red, etc.). Image schemata are structures that are constitutive of our common bodily experience. Compared with basic-level concepts, they are a lot less structured, they exhibit structure only along general lines. Typical examples are the CONTAINER schema (structural elements: an interior, a boundary, an exterior), the SOURCE-PATH-GOAL schema (structural elements: a starting point, an end point, a path from starting to end point, direction toward the end point) the LINK schema (structural elements: two entities, A and B, and a connecting link), the PART-WHOLE schema (structural elements: a whole, parts, and a configuration), and the UP-DOWN schema. These pre-conceptual structures of experience are directly meaningful, not least because they are immediately and constantly experienced arising from the way our body functions in an environment.

Because we also apply abstract concepts, the question arises as to how abstract concepts can develop from basic-level and image-schematic structure. For one thing, by metaphorical projection from the direct experienced phenomenal physical domain to abstract domains; for

another thing, by projection from basic-level to super- and subordinate categories. We will restrict ourselves to an explication of the former. The following examples of metaphorical projection are typical: We understand the target domain QUANTITY in terms of structures of the source domain VERTICALITY (more corresponds to up, less to down: the VERTICALITY metaphor makes QUANTITY intelligible); arguments as a target domain are made intelligible in terms of structures of the SOURCE-PATH-GOAL schema (the first premise as starting point, the conclusion as end point, etc.); we understand slavery as bondage or freedom as not being tied down (the LINK schema as the source domain of the metaphorical projection); logical structures like transitivity ($A \rightarrow B, B \rightarrow C, \text{ hence } A \rightarrow C$) are understood within the frame of the CONTAINER schema (B contains A and C contains B; hence, C contains A).

What the last example (more importantly) should make clear is that the logical structure (in the example, transitivity) is already meaningful to human beings due to their bodily experiences. It is just not the case that disembodied, meaningless logical structures form the scaffold of thought. In fact, the structure is already meaningful and tailor-made for the human body. Hence, understanding and thought are not sensitive only to meaningless structures.

According to the spatialization of form hypothesis, the spatial structure (to which several image schemata correspond) is mapped onto the conceptual structure: For example, categories are understood in terms of CONTAINER schemas, relational structures are intelligible in terms of LINK schemas, or radial structures are taken in terms of CENTER-PERIPHERY schemas. If one applies the theory to itself, that is, if one asks about the way metaphorical projections are understood within cognitive systems, then it seems unproblematic to say that they are also understood in terms of projected image schemata: Within the frame of the CONTAINER schema, source and target domains are understood as being set off from one another, the metaphorical mapping from source to target domain (the conceptual domain) is understood in terms of the SOURCE-PATH-GOAL schema, although the path as a structural element is unspecified.

If the cognitivist objectivist semantics takes the following principles for granted, it neglects how the human conceptual system (in contrast to artificially construed systems) works: (1) Every concept is either primitive or composed of primitives by completely productive principles of semantic composition. (2) The internal structure of conceptual systems arises entirely from the application of completely productive principles of semantic composition. (3) Only those concepts without internal structure are directly meaningful. Both basic-level and image-schematic concepts have internal structure and are directly meaningful – hence at

odds with (3). Additionally, basic-level and image-schematic concepts have internal structures that cannot result from applying completely productive principles of composition – hence at odds with (2). (Lakoff 1987, pp. 269-303)

In connection with the phenomenally catchy notion of image schemata the question arises whether this grounding of conceptual processing in preconceptual sensorimotor structures is also neurally plausible. The following section is supposed to show that it is really like that.

1.3 Neural exploitation – how concrete and abstract concepts make use of the sensorimotor system (Gallese & Lakoff 2005)

The following deliberations on neural exploitation anticipate the basic thought of PSS (see Chapter four), namely, that higher-level cognition bases on the re-enactment of sensorimotor areas. The continuity involved between the representational formats of online sensorimotor and cognitive processing is a keynote of this work.

Understanding concrete concepts (phenomenally re-presented physical actions and objects) requires sensorimotor simulation. Sensorimotor simulation is carried out by the sensorimotor system of the brain. What one understands of a sentence in a certain context via sensorimotor simulation is the content of that sentence in this context. This is incompatible with the claim that concrete concepts are modality-neutral and disembodied.

Since any theory of concrete concepts must capture all of the existing sensorimotor structure (agent-object-location, manner, purposes, and phases), a modality-neutral theory of concepts needs to assume that this structure is represented neurally outside the sensorimotor system. Because this sensorimotor structure exists in the sensorimotor system, it would have to be duplicated outside the sensorimotor system in order that such a modality-neutral theory could be true. Proponents of an embodied neural theory of concepts simply appeal to an Occam's Razor argument – the duplication is just superfluous.

Basic-level concepts are constituted by the convergence of the gestalt perception of objects (both observed or imagined) and the motor programs which account for interaction with objects (again, both performed or imagined). Functional neural clusters bring the perceptual and motor properties together (sensorimotor coupling).

Taking the example of the action concept *grasp*, I will now elucidate three central points: that single neurons fire 'multimodally', that 'functional clusters' realize multimodal firing, and that understanding is 'mental simulation'.

To say that an action such as grasping is multimodal here means that its neural enactment consists in the activation of neural substrates which are used for both perception and action. It has been demonstrated, for example, that the premotor area F4 comprises neurons that integrate motor, visual, and somatosensory modalities in order to control actions in space and to perceive the space reachable by body parts ('peri-personal space') (e.g., Rizzolatti & Gallese 2004). Moreover, it has been shown that F4 neurons also integrate auditory information about the location of objects within peri-personal space (Graziano et al. 1999). What is critical is that the same neurons that control target-oriented actions also respond to visual, somatosensory, and auditory information about objects to which they are directed. They respond that way because they are part of a certain cortical network – in other words, they are part of a certain functional cluster.

What is a network in this context? In addition to the two other parietal-premotor networks that realise multimodal functions, the F4-VIP functional cluster transforms the spatial position of objects in the peri-personal space into the most apt motor programs for interacting with these objects. If this functional cluster is damaged, conscious awareness of, and interaction with objects localised therein are not possible (Rizzolatti et al. 2000). Here the spatial position of the objects is much more important than their properties. The F5ab-AIP functional cluster comprises 'canonical neurons' that transform physical features of objects (such as shape, or size) into the most apt hand-motor programs for acting on them. Whereby the properties of the objects are much more important than their spatial positions. Damage to this cluster would result in visuomotor grasping deficits, whereby motor capacities for grasping would remain intact (e.g., Fogassi et al. 2001). The F5c-PF functional cluster comprises 'mirror neurons' that respond when subjects perform goal-related hand actions *and* when they observe other individuals performing similar actions (Rizzolatti 2001 et al.). Where does mental simulation come into play?

Note that the same neurons that fire when a monkey turns its head toward a certain location in its peri-personal space also discharge when an object is in place, or it sounds, at the same location toward the monkey would turn its head, if he would actually do. Thus, action simulation is responsible for automatically triggering an action plan in response to the sight or sound of an object at the respective location. If the neurons fire in the presence of the sight or sound of the object without turning the head, it makes sense to say that these neurons simulate the action toward the respective location. This takes place within the F4-VIP cluster. Within the F5ab-AIP cluster are 'canonical neurons' that fire both when the monkey actually is grasping an object and when it sees an object that it could grasp but does not. It is indeed

the case that the same neurons that respond for the execution of a certain manner of grasping a particular object also fire if that same object is merely observed. Observing a graspable object only triggers those neurons that provide a suitable manner for interacting with it (Gallese 2003). Obviously, if the motor programs are triggered without being executed, one speaks of action simulation.

The 'mirror neurons' within the F5c-PF cluster do not discharge if the object one could act on is simply seen. They also do not respond if the observed action is performed with a tool. Indeed, some mirror neurons are (about 30%) 'strictly congruent' and fire when the action seen is exactly the same as the action performed. Others (about 70%) fire when the monkey either grasps with a pincer grip or perceives any type of grasping. The point is that when a monkey sees or hears another subject performing an action it simply simulates the same action.

What is the corresponding evidence for embodied simulations in the case of humans? Firstly, homologous to the F4-VIP cluster in monkeys, a functional cluster that fires when stimuli in the peri-personal space are seen or heard was found in humans. What is critical is that a premotor area is activated during such perception, an area that controls movements in the peri-personal space (Bremmer et al. 2001). Secondly, it was shown that while observing, silently naming, and imagining using man-made objects, the subjects' ventral premotor cortex was active – an area that discharges if subjects are using the respective tools to perform actions (e.g., Chao & Martin 2000). These neurons stand for the 'canonical neurons' found in monkeys. Thirdly, several brain-imaging studies show that while humans are observing actions of conspecifics, premotor and parietal areas are active which most likely are the homologue to the mirror system found in monkeys (e.g., Buccino et al. 2001). The mirror system matches action observation and execution.

Because these types of simulations are triggered by external input, they can be regarded as more or less online processes. That is not to say that the simulated contents are only determined in terms of bottom-up information, it quite plausible to assume that memorized information is relevant to those simulations. Moreover, are there simulated contents that can be internally triggered, in addition to that they bear on memorized contents?³

Mental imagery as embodied simulation embraces both embodied visual imagery and embodied motor imagery. Visual imagery makes use of some of the same neural areas that

³ This internal triggering is understood as being internal to the effect that the simulated contents bear no perceptual resemblance to the triggering inputs. Concretely, if you read a sentence that triggers a perceptual simulation whose contents have nothing to do with the phenomenal environment in which you are located, then I take this triggering as being internal. I think of lack of perceptual resemblance as a necessary condition for internal triggering.

are activated during seeing (Kosslyn 2006, Farah 2000). It was shown, for instance, that the time needed to scan a visual scene is quasi identical to the time it takes to scan the scene in imagination (Kosslyn et al. 1978). It was further probed that imaging and actually perceiving the same scene share certain neural correlates (Farah 1989, Kosslyn 1994), such as the primary visual cortex.

During motor imagery, some of the same brain areas are used which are responsible for action. Brain-imaging studies have demonstrated that executed actions and motor imagery both activate a common neural network – consisting of the premotor cortex, the supplementary motor area (SMA), the basal ganglia, and the cerebellum (cf. Jeannerod 1994). Evidence for the embodied nature of motor imagery is that heartbeat and breathing frequency increases when one simulates bodily performance (Decety 1991).

Complex premotor structures are called 'executing schemas' (or X-schemas). Actions differ from each other in the kind of premotor structure they make use of and how they are linked to the motor cortex and other sensory areas that subserve perceptual and somatosensory feedback. What is most important for our purposes is that the X-schemas of the premotor system can function independently of the execution of motor behaviour – in understanding language or planning action, for example.⁴ In respect thereof, this comes close to Clark and Grush's concept of 'full-blooded representation' (1999, p. 10), because the independent functioning of the premotor cortex is a mechanism that can be fully decoupled from ongoing environmental input. The same applies to visual imagery.

1.4 Sensorimotor coupling and a weakened version of the dual visual systems hypothesis

As sensorimotor coupling was an essential part of the last section, the question arises whether it is compatible with a meanwhile known thesis, namely, the dual visual system hypothesis. Prima facie its basic idea, that is, a functional distinction between vision-for-action and vision-for-phenomenal-perception, discounts sensorimotor coupling.

⁴ This question is obvious: How are these neural mechanisms related to basic-level concepts and image-schemata? The sentence *The cat is on the mat* is true only relative to *our* embodied understanding of the world – basic-level categories and words for them like *cat* and *mat* are recognizable by means of our gestalt perception, and we can interact with them in virtue of our motor programs (that is, the different X-schemas). ON can be analysed in terms of three kinesthetic schemas: ABOVE, CONTACT, and SUPPORT. The category animal neither directly corresponds to our gestalt perception, nor do we have motor programs for interacting with animals in general. The homologue of canonical neurons in humans could make up the neural correlate of basic-level categories.

The dual visual systems hypothesis of Milner/Goodale (1995, 2005) is characterized by these three theses: Vision-for-action and vision-for-perception are functionally distinguished, this functional distinction is reflected by the anatomical segregation between the dorsal and ventral stream of human vision, and awareness is restricted to vision-for-perception. This functional distinction is *inter alia* confirmed by the fact that people are not under certain phenomenal visual illusions (e.g., the Ebbinghaus illusion) if their behaviour is controlled by unconscious vision (Milner & Goodale 1995, Chapter six).

According to a weakened version of the dual visual systems hypothesis (which avoids a strict isolation of these two streams), the degree of stream independence and the nature and extent of stream interaction are task- and attention-dependent (cf. Jeannerod 1997; Decety & Grezes 1999). Moreover, Pascual-Leone and Walsh (2001) showed that feedback from high (V5/MT) to low-level visual (V1 and V2) areas is necessary for certain forms of conscious visual perception. Nevertheless, the crucial insight remains – fine-action-guiding visual processing is often carried out independently of the processes underlying phenomenal vision. Clearly, sensorimotor coupling can take place both at the non-phenomenal and at the phenomenal level. Thus it is well compatible with the two visual systems hypothesis.

The following section draws a distinction that is in line with setting vision-for-phenomenal-perception apart from vision-for-action.

1.5 Body schema and body image – double dissociated phenomena

A body schema is a system of sensorimotor processes which permanently controls posture and movement and thereby functions without reflective awareness and conscious perceptual monitoring. Body schemas can also be conceived of as an assemblage of sensorimotor interactions which define a specific movement or posture, such as the rotation of the ankle within the larger movement of curling a free-kick. The body image, by contrast, is a sometimes conscious system of perceptions, attitudes, beliefs, and dispositions with reference to one's own body. It involves at least three different aspects: body percept, body concept, and body affect. The body percept is the subject's perceptual experience of his or her own body; the body concept is the subject's conceptual understanding of the body in general; and the body affect is the emotional attitude of the subject toward his or her own body. Whereas the body image phenomenally depicts the body as clearly differentiated from its environment, the body schema functions in an integrated way with its environment.

In a case of unilateral neglect, the patient has an intact body schema but a missing body image. Stroke patients sometimes do not perceive or attend to one side of their body. This is the result of brain lesions in the hemisphere opposite the neglected side (contralateral). As excluded from the body image, the respective side is ignored, denied, and sometimes disowned, as if it does not belong to the patient. Decisively, though those patients are not visually conscious of the body parts of the neglected side, they are able to use this side to dress, walk, eat, and so on. These complex spatial and motor skills are explained by an intact body schema.

In instances of deafferentation, the contrary is the case: Though the patient (IW) has an intact body image, he lacks an intact body schema. If you ask the patient to close his eyes, point to his thighs, and if you move his thighs, he is unable to point to his thighs, since without phenomenal vision and proprioception (due to a lacking body schema), he does not know where his hand and his thighs are located. A person with normal proprioception has no problem with this task. The patient has no sense of touch and proprioception below the neck as a result of a sensory neuropathy in which large fibres below the neck have been harmed due to illness.

The body-schema system can be considered as satisfying three functions: (1) processing new information about posture and movement; (2) producing movement patterns; (3) enabling intermodal communication between proprioception and other modalities.

(1) Visual proprioception and visual kinesthesia (cf. Gibson 1979, Neisser 1976) provide information about the environment and the way the organism moves through it is directly related to the body schema. Supported by the vestibular system, these two sources of information allow one to distinguish between movements made by environmental objects and one's own movements. Non-visual proprioception, in the sense of somatic information about joint position and limb extension, is the major source of information of present bodily position and posture. The pre-reflective proprioceptive awareness (feeling one's thighs with one's eyes closed) has to be distinguished from non-conscious proprioceptive content processing – thus, somatic proprioception is twofold. (2) At the behavioural level, a motor schema corresponds to an elemental aspect (e.g., rotating one's ankle) of a larger movement (curling a free-kick). At the neural level, a motor schema corresponds to the neuronal activity required for this elemental aspect of the whole movement. Visually-guided movements usually activate the dorsal stream and feed into the system for initiating a motor program. IW's visual control of movement, in comparison, activates the ventral stream and orbito-frontal cortex (areas responsible for cognitive and non-motor visual tasks). One way to

explain IW's lack of proprioceptive information about motor programs is that proprioception fails to note the present motor state and therefore the motor program cannot be accessed in that way (e.g., he cannot easily hold a conversation while he is dressing, since the latter needs his full attention). (3) Normally there is an intermodal translation between vision and proprioception that allows visual perception to inform and coordinate behaviour and proprioceptive simulations of another person's movements – an innate feature of the human sensorimotor system. The patient IW can compensate for the deficits by processing the relevant information about his own body via the body image.

Since IW's sense of ownership was re-established very quickly, and we normally have a sense of ownership with closed eyes, the body image and the body schema play an important role in constituting our sense of ownership. The fact that IW's control over movements and his sense of agency were regained (cf. TMS study in Cole 1995) by using aspects of his body image shows that the sense of agency is not tied to proprioceptive feedback. Moreover, the more one can make one's movements automatic by using learned motor programs, and the less attention such movements thus take, the more holistically embodied they appear.⁵

1.6 At a subpersonal functional (non-neural) level – the inverse model and the forward model

Also due to the fact that I will later draw on the forward model to reply to an objection to concept empiricism (namely, that solving the frame problem does not come for free for theories that ground higher-level cognition in the sensorimotor systems), I am now going to introduce the notions of 'forward model' and 'inverse model'.

The inverse model functions in a simple adaptive control system for general purpose motor control. This control system can be understood analogously to a thermostat consisting of six elements: (1) A target signal (the desired room temperature); (2) an input signal (the actual room temperature) that is the joint result of (3), exogenous environmental events and the output of the control system (heat output); (4) a comparator that determines whether target and input signals match and the degree of the match/mismatch (e.g., the room is seven degrees below the desired temperature); (5) output that is determined by comparing target

⁵ Furthermore, the three notions of motor space, proprioceptive space, and conscious perceptual space have to be distinguished. Motor space and phenomenal perceptual space differ in that the former functions in an egocentric space of reference and the latter in an allocentric frame of reference. Mirror drawing also shows how proprioceptive directions come into conflict with phenomenal visual directions. (Gallagher 2005, first two chapters)

and input signals (e.g., decreasing the heat output); (6) a feedback loop by which the output effects the succeeding input signal (e.g., actual room temperature decreases). The function which maps target signals to the output in the context of actual input signals is called an inverse model. The feedback which operates in real time is designated as re-afferent feedback – that is, input to a system resulting from the organism's own activity (by contrast with ex-afferent input which results from exogenous events). Visual and proprioceptive inputs in consequence of moving one's limbs or movement through space are re-afferent inputs. This system is adaptive in that it adjusts itself to altering conditions in the environment and compensates for exogenous disturbances. The control process is dynamic and cyclical – information about inputs is not cut off from information about outputs; and the control process does not have discrete steps or a non-arbitrary start or finish.

The idea of a forward model consists in mapping the output signal back onto the input signal (what is often called 'efference copy'). The function of the forward model is to predict the consequences of the output on input. Efference copies generate a simulation of the expected effects of the output which allows the speeding up of the control process and the smoothing of the respective behavioural trajectory. This simulation is low-level, because it can perform its speeding and smoothing without being globally available, whether it uses actual or simulated feedback. (e.g., Wolpert et al. 2003) Comparing efference copy with re-afference provides the required information to distinguish between self-activity and activity of the world. Cell assemblies which mediate the connection between efference copy and input signals could have both motor and sensory fields. Suppose that you grasp an apple and bite into it, then there will be an association between your efference copy for the grasping and eating movements and multimodal inputs characteristic of such movements. Canonical neurons are good candidates for such sensorimotor affordance neurons. (Hurley 2008)

In the next I would like to demarcate our understanding of embodied and embedded mind from a position that is called 'extended functionalism'.

1.7 Embodied and embedded mind without extended functionalism

The theses of both embodied and embedded mind must be distinguished from the thesis of extended mind. According to the latter some mental processes are partially *constituted* by processes of environmental manipulation. Since mental states and environmental structures satisfy the same (coarsely individuated) functional roles, they are identical. These functional roles are taken as being multiple realisable – they are either realised by neural or non-neural

external structures. (Wheeler forthcoming, pp. 3-5) By contrast, the thesis of embedded mind only speaks of a dependence of mental processes on the environment – that is, some mental processes only function in connection with environmental structures. In other words, being situated in a wider system of scaffolding, mental processes are vitally facilitated. Since the thesis of embodied mind only says that some mental processes are partially constituted by non-neural bodily structures and processes, it is obvious that it must be separated from the thesis of extended mind. (Rowlands 2009, p. 54) As we will see later in more detail, the crucial problems of extended functionalism is an too undifferentiated view of multiple realisability and the neglect of neurally embodied conceptual structures.

That which advocates of extended functionalism consider as external is regarded in the following chapter as belonging to a neurally generated phenomenal world model. However, that is not to say that extended functionalism is true within the scope of our phenomenal world model. Moreover, Lakoff and Johnson's idea that conceptual understanding is grounded in sensorimotor structures will be revisited in the second section of the following chapter. What's more, the concept of simulation will be considered in respect of its epistemological status.

2 Conceptual tools: Representation, simulation, and presentation

Presentata, through their output decoupling, enable the system to develop a larger behavioral repertoire relative to a given stimulus situation. Representata integrate those basic forms of sensory-driven content into full-blown models of the current state of the external world. Advanced representata, through input decoupling, then allow a system to develop a larger *inner* behavioral repertoire, if they are activated by internal causes – that is, as *simulata*. (Metzinger 2003, p. 49)

The overall purpose of this chapter is to provide an exhaustive theoretical framework (including a uniform terminology) for mental content. The distinctions included therein are especially relevant to the Chapters four and six, because concepts like intentionality, reference, and truth cannot be grasped adequately without having a precise and comprehensive typology of mental occurrences. This typology is mainly borrowed from Thomas Metzinger's *Being No One* (2003).

2.1 A *mind-dependent reality* – conscious experience as a virtual world

Given that a naïve realistic understanding of phenomenal visual perceptions is subject of my criticism of classical theories of mental content, and given that, as previously mentioned, higher-level cognition is grounded in phenomenal representations and we will appeal to PSS as a theory of higher-level cognition later (see 4.3), it is not of no significance to clarify the epistemological status of phenomenal representations.

What is meant by reality? Everyday objects, such as tables, cars, or trees are mind-dependent. Irrespective of whether one speaks (possibly in the frame of philosophical theories) of them as ontological entities or in terms statements about those entities, they are nothing more than objects of our phenomenal world model created by the brain. With respect to those entities, the distinction between a 'minimal' and 'qualitative realism' is not particularly relevant, because their existence as basic elements of reality is already mind-dependent. This is a position that in the classical realism/anti-realism debate is called 'eliminative anti-realism' (Willaschek 2003, p. 12).

What is meant by dependent? Does the classical distinction between causal dependency (an event A is causally dependent on another event B iff B belongs to the causes of A) and conceptual dependency of thought (the statement that p is conceptually dependent on the statement that q iff q is non-exclusively logically implied by p and q is a statement about mental occurrences) (cf. *ibid.*, pp. 29-34)⁶ work? An 'extensional' dubbed model of causal relations between ordinary objects is as phenomenally, and therefore neurally, constrained as our statements about mental processes. To act as if the whole problem of mind-world relation is nothing more than explaining how we can form reliable beliefs about phenomenally perceived objects falls significantly short, particularly as the seen table is not the cause of your phenomenal visual perception but the result of your brain's construction process. Any distinction between extensional (at least with respect to everyday objects) and intensional sentences takes place against the background of our phenomenal world model. To this extent, the phenomenally represented world is already mind-dependent, and in a causal way.

Last but not least, what is mind – conceived in quite general terms? The dynamical self-organisation of the brain, which, *inter alia*, generates processing on the level of global availability. More about this later.

⁶ For example, when somebody kicks against a door and as a result the door begins to vibrate, then the vibration is causally dependent on the kicking. The sentence, "It is more reasonable not to protest against it", is conceptually dependent on thought, because it implies that some people *regard* the renunciation of protest as more reasonable and this is a statement about mental occurrences.

Given that our physical body interacts with the physical world and that our phenomenal experiences function, *inter alia*, to enable complex behaviour, and given that phenomenal structures can be processed unconsciously within the organism also to coordinate behaviour in the physical world, phenomenal representations have to preserve some structures of the physical world.

Considering that it is a strong, far-reaching, and, for our purposes, not unimportant thesis that our phenomenal perceptual world is neurally constructed, what reasons suggest that this is true? Firstly, the natural function of mental content is to generate intentional representations. In order to be behaviourally related to the environment, any biological agent must exchange information with the environment. On the physical level of description, the biological agents are exposed to diverse energies – electromagnetic, mechanical, and chemical energy. Only because of interaction with the living organisms can the energy be determined in terms of stimuli – that is, visual or auditory stimuli, for example. By virtue of interaction between organisms and energy, the resulting stimuli are transduced into a 'common informational code': The receptors of the different sensory modalities convert the different types of energy into action potentials (electrochemical excitability of cells), that is, the common code. This common code enables communication between billions of neurons of the central nervous system. (Gallese 2003, p. 1232) The components of our visual system process the information contained in photons that are part of the external world in order to make location, texture, movement, colour of environmental objects accessible for the mind. This is similar in the case of the auditory system, that uses information contained in atmospheric compression waves to determine the spatial position of objects. (Hardcastle 1999, p. 107) Secondly, as an argument against the justification of claiming that vision reconstructs the physical world, we ought to be able to compare the physical objects with our visual representations so as to decide whether a physical world is reconstructed. That, however, is excluded as a matter of principle: Even if we use high-tech instruments to extend the range of our senses, our visual system always has to generate a phenomenal representation. (Davies et al. 2002, pp. 79-81) If one admits phenomenal theoretical models of the the physical world as justificatory contrast, then to assume a strong reconstructive nature is wrong, given the above-mentioned first point. Thirdly, evidence from quantum mechanics suggests that the psychological spatial system (instantiated in networks of neurons – e.g., the hippocampus as the allocentric spatial mapping system), in which discrete entities causally interact with each other, does not correspond to the physical reality. For example, properties like position and velocity, which are independent of each other at the macroscopic level, are not so at the microscopic level.

(cf. O'Keefe 1993, pp. 47-51) Fourthly, clearly, from the fact that one can bring forth phenomenal presentations by electrically stimulating certain representational brain areas it does not follow that in normal circumstances our phenomenally perceived environment has no structural overlap with the physical world. What this at least entails is that the structures of our phenomenal world are entirely internally constructable. Fifthly, the fact that neuroscientists have determined a wealth of neural areas respectively necessary for certain phenomena or capacities (e.g., motion, depth, form, colour) and given that those neurophenomenological structures are the product of evolution, to which other animals with another neurophenomenological structures also belong, suggests that our phenomenal experiences do not have direct access to objective reality. Sixthly, there is no one-to-one correspondence between physical properties and phenomenal experiences – the same green²⁷ is caused by different mixtures of wave lengths. (Metzinger 2009, p. 20) Seventhly, due to the fact that the transduction and conduction velocities of different sensory modalities differ from one another, the system itself needs a 'window of simultaneity' in which multimodal object representations occur (e.g., we experience the taste and colour of a red apple at the same). Thus, this temporal presence does not literally take place in physical systems, but is phenomenally generated. For example, it seems empirically plausible that elementary sensory information (e.g., colours, motion properties, shapes) is integrated into the conscious experience of a multimodal object by the synchronization of neural responses (Singer 2000). If these points are not considered as sufficient reasons for the existence of a phenomenal world model, I take its existence at least for granted with respect to my further argumentation.

2.2 Mental and phenomenal representation

Beforehand it is important to say that the subsequently presented concept of representation differs from Clark and Grush's notion of 'full-blooded representation' (1999) (see 1.3). The central functional property of 'full-blooded representation' (that is, internal structures that can reactivate multimodal experiences independent of ongoing environmental input) is taken into account by our concept of simulation rather than by our or Metzinger's concept of representation.

The first concept I explicate is mental representation. Mental representation is analysed as a three-place relationship: the representandum as the object of the representation; the representatum as the concrete internal state which carries information concerning the object;

the representation as the process by which the system as a whole produces the internal state.⁷ In the case of the representatum, it is important to make sure that one does not commit the 'error of phenomenological reification' (Metzinger 2003, p. 22): What one is likely to experience as a stable content is itself constituted by an ongoing process that phenomenally erases its own temporality. Here, the phenomenal experience itself suggests the neglect of the transition from mental processes to stable mental objects. Moreover, the precise way we refer in natural languages to phenomenal contents erases the dynamics of phenomenal and informational processing. When we speak of *a* content of *a single* phenomenal propositional attitude, the experiential content of an ongoing representational process is reified. This reification brings with it the danger of committing the classical phenomenological fallacy: Clearly, the content of qualitative experiences such as a hallucination of a three-armed pink elephant, for example, cannot be analysed as a non-physical object that, inter alia, possesses the property 'pinkness'. The above-mentioned phenomenal now is a virtual actuality that is probably a teleofunctionalist property of the biological system: The representatum represents a part of the world (both external and internal) *for* a certain biological system in order to achieve its aims.

Now let us shift our attention to the notion of phenomenal representation. A functional criterion for demarcating phenomenal representations is that they carry exactly that information which possesses the following three dispositional properties: availability for guided attention, for cognitive processing (e.g., availability for generating concepts), and availability for behavioral control. It is useful to distinguish between four kinds of introspection: (1) Introspection as 'external attention' is a subsymbolic meta-representation that operates on an internally generated world-model (thereby serving teleofunctional purposes) and takes its intentional content as external. It corresponds to the phenomenal experience of attending to environmental objects. (2) Introspection as 'consciously experienced cognitive reference' is a conceptual meta-representation that phenomenally represents the cognitive reference to an internal state whose content itself is regarded as external. The cognitive reference itself is phenomenally experienced (it operates on a world-model as well). (3) Introspection as 'inward attention' is a subsymbolic meta-representation that operates on an internally generated self-model. Thus the introspective experience focuses

⁷ This is not taken to mean that the representation ascribes properties to the representandum (to be understood against the backdrop of our phenomenality). The way in which the representandum is represented most often entails the information which is to be carried by the representatum. One could say that the something is represented as existent.

Clearly, if one makes the assumption that our physical body exists in the physical world, then the brain must represent in a way which allows the more or less successful existence in the physical world.

on an internal state whose content is regarded as internal. (4) Introspection as 'consciously experienced cognitive self-reference' is a conceptual form of self-knowledge directed toward internal states that are internally regarded as internal (it operates again on a self-model). We consciously experience that we are currently cognitively referring to our own states.⁸

What does it mean that conscious experience is *subjective* experience? Functionally, information is integrated into an internal model of reality of an individual system which thereby gets privileged introspective access to this information. At the phenomenal level, subjective information can be integrated into the conscious self-representation the system is currently having. Introspection as 'inward attention' and as 'consciously experienced cognitive self-reference' are those processes which make information subjective in a phenomenal sense.

Availability for cognitive processing can be characterized by the following principle: A necessary condition for becoming the content of cognitive reference is that the content is phenomenally re-presented – what is called the 'principle of phenomenal reference'. Phenomenally re-presented information can be categorized and memorized, it is 'recognizable' information. Whereas processes like day-dreaming or low-level attention may be initiated by unconscious information causally functioning within the system, self-triggered, explicit cognition exclusively operates on phenomenally re-presented information. Availability for the control of action with respect to phenomenally re-presented information is limited to a certain class of actions – namely, '*selective* actions' directed toward the phenomenally re-presented content. This kind of availability is clearly reliant on sensorimotor integration: In order that phenomenally re-presented contents are capable of controlling action, they have to be directly fed into the mechanism that activates motor representata. Since basic actions are physical actions or bodily motions, we need an internal representation of the body. Thus the functional role of the phenomenally re-represented content is that it can be directly fed and integrated into a dynamical representation of one's own body as a presently acting and ongoing acting agent. This agent, however, has the flexible capability of swiftly decoupling motor and sensory information processing (autonomy).

To conclude these remarks on representation, three forms of representations must be distinguished: *Internal* representations are structures in the brain that fulfil a function for the system as a whole, that admittedly possess certain contents (e.g., action-relevant information

⁸ I take the concept of symbolic representations as those internal components that can be processed in the absence of externally triggered phenomenal representations which could be their contents.

processed in the dorsal stream or proprioceptive information of the body schema)⁹, but are never conscious or have phenomenal content. *Mental* representations have the dispositional property of becoming available for attention, cognitive processing, and control of action. They carry certain contents but do not have phenomenal content, if certain additional criteria are not fulfilled. *Phenomenal* representations have to meet a number of different functional constraints. Two of them were already stated: temporal internality and global availability.¹⁰

2.3 Mental and phenomenal simulation

Mental simulation is the third concept we are dealing with. Mental simulation is a three-place relation composed of an individual content processing system for which a counterfactual situation is simulated by a physically internal representatum. The intentional content of the counterfactual situation *can become available* as representandum for subsymbolic forms of introspective attention, for symbolic forms of cognitive reference and for selective control of action. Interestingly, from a physical and epistemological point of view, phenomenal representations and simulations coincide – due solely to the fact that our brain creates a temporal frame of reference, we never have direct epistemic access to the surrounding world. Since the content, represented as phenomenally actual and objectively real, is not actual and objectively real, but internally defined that way for teleofunctional reasons (virtually), it can be considered as simulational, just like simulations of counterfactual situations. Representata and simulata differ only on the phenomenological level of analysis. The internal states taking part in simulational processes can admittedly be elicited by external stimuli, but are not stimuli-correlated in the strict sense. Typical examples are thoughts of things that are not sensorially present during routine activities, for instance, thinking about a TV show while driving a car. The intentional contents of mental simulations can become available for subsymbolic forms of introspective attention, for symbolic cognitive reference, and can become globally available for selective control of action.

Now let us consider the phenomenal version of simulation. All states of affairs we can consciously simulate (imagine or conceive of) possess the property of being phenomenally possible. Clearly, what is phenomenally possible depends on the layout of our consciousness, which to a great extent can be characterized functionally. In addition to the notion of possibility, phenomenal simulations are characterized by a transparent representation of

⁹ These examples at least express my interpretation of ‘internal representation’. Exceptionally, ‘internal’ is hereby used to mean those inner episodes that cannot become globally available.

¹⁰ In the further discourse of the work, I will not go into the other constraints.

temporal internality – the now is internally construed as real. At least as important as is the fact that phenomenal simulations always take place against the background of a phenomenal sense of ownership. They are experienced as belonging to a subject that is viewed as real. One function of phenomenal simulations is to produce world models that are biologically relevant, since they allow the planning of goal-directed actions, for example. To make this function possible, a representational frame is required that serves as 'evaluative context' for the simulational contents. This evaluative context is a world-model that is defined as actual for the system ('world zero hypothesis'). If there were not such an internally defined 'reference model', phenomenal simulations would be experienced as currently real, and that would make important functions like future planning impossible.

In the following section I am going to elucidate the concept of presentation. These remarks are especially relevant to the theory of nonconceptual mental content that is constructed in the fourth section of Chapter six.

2.4 Mental and phenomenal presentation

Now let us focus attention to the concept of *mental presentation*. To realize what is meant by mental presentation, blindsight patients are a good example. They are capable of discriminating diverse colour stimuli by means of predicates like “blue” or “green” in normal perceptual contexts without any accompanying colour experience (cf. Weiskrantz 2009).¹¹ That ability is based on their comparatively normal sensitivity for different wavelengths within the scotoma. This example is said to clarify that mental and phenomenal presentation differ at least in terms of the degree in which their stimulus information is globally available – although, for example, the content of phenomenal presentations is available for selective action, mental presentational information in the case of blindsight is only available for a form of motor selection.¹²

The last concept I delineate is *phenomenal presentation*. Phenomenal presentations can be characterized by four different principles: 'the principle of presentationality', 'the principle of reality generation', 'the principle of nonintrinsicity and context sensitivity', and the 'principle of object formation'. Phenomenologically, the principle of presentationality states that presentation is subjectively experienced present in the context of sensation – for

¹¹ It is believed that the discriminations are so much fine-grained that they cannot be considered as representations.

¹² Furthermore, there is empirical evidence that the perceptual processing of non-phenomenal stimuli fulfils two significant functions: For one thing, it biases what is phenomenally experienced; for another thing, it influences how stimuli are consciously experienced (Merikle et al. 2001).

example, *green₁₇* is always experienced as *green_{17-now}*. Clearly, it is integrated into the higher-order phenomenal re-presentation of time. Normally, if it is the most simple form of phenomenal content, we are not able to deliberately imagine or remember it. Functionally, different forms of presentational content constitute the phenomenal frame of reference (that is world₀). To have such a frame of reference allows us to separate phenomenally the present reality from mental simulations of counterfactual situations. On one hand, the presentational content as nonconceptual content points to a specific sensational feature, on the other, it invariably points to the fact that this feature in the environment or the organism's own body is given *de nunc*. The principle of reality generation stands for the subcognitive level of phenomenal presentations which almost inevitably suggests the existence of whatever is currently presented to us. Both the sensory now and the presentational appearance of objective existence have a virtual character that is generated by subpersonal processes which take time. The principle of nonintrinsicity and context sensitivity repudiates the well-known philosophical notion that phenomenal properties are essential properties in the sense of being the non-relational, context-invariant essence that becomes exemplified by the contents of simple sensory experiences (Levine 1995). It has been shown, for example, that a homogeneous field of a single coloured light appeared neutral instead of coloured just as the perceptual context of the preceding visual scene disappeared.¹³ The principle of object formation stands for pre-attentional processes of feature integration bringing about the experience of coherent perceptual gestalts.

In summary, it can be said that six different forms of content function within human beings as biological systems. Concerning higher-level cognition, the phenomenal kinds of content are the most relevant. Our concepts are only formed against the background of phenomenally accessible contents. Among these, only re-presentationally and simulationally processed contents have the dispositional property to be available for cognitive processing in the form of concepts. Of course, that is not to say that concepts once extracted from phenomenal representations cannot function non-phenomenally. It seems to be quite possible that non-phenomenally processed content is directly fed into speech behaviour without being accompanied by phenomenal simulation.¹⁴

¹³ I do not explicitly argue for this, but there is a lot of empirical evidence that presentational content is a relational phenomenon which depends on the existence of a perceptual context and that it supervenes on complex causal relations and therefore is by no means capable of existing by itself across such contexts (cf. Metzinger 2003, pp. 100-4).

¹⁴ If one wanted to formulate a theory of thinking, it seems to me that speech supplied with non-phenomenally processed content would be a part of it.

2.5 Phenomenal mental models

What is a phenomenal mental model? Functionally, phenomenal mental models are globally available for cognition, attention, and the control of behaviour. Because phenomenal mental models consist of supramodally experienced objects (e.g., seeing and feeling an apple simultaneously), they are available for the formation of mental categories and concepts. Any phenomenal mental model has to be integrated into the dynamic process of modelling the current presence. Since the contents of subjective experiences are the contents of a mental model of the world as a whole, phenomenal mental models need to be currently integrated into the present mental world. Moreover, phenomenal mental world models contribute to the emergence of the first-person perspective: If the system phenomenally models the self-world boundary (the phenomenal self-model (PSM)) and the ongoing subject-object relations (the phenomenal model of the intentionality relation itself (PMIR)), the experience emerges of a subject that interacts purposefully in the world. Most phenomenal mental models are transparently represented – the fact that we are actually modelling a reality is not globally available for attentional processing. An exception are reasoning or planning processes that are phenomenally represented as internally generated representations. In other words, they are based on opaquely represented phenomenal mental models. Phenomenally representing inferential relations between language-like propositions can be understood as a phenomenal mental model in which the form/content distinction is reintroduced on the level of phenomenal content.

2.6 Thoughts on the vehicle-content distinction

What we need is *embodied* content, as it were – an ongoing and physically realised process of *containing*, not 'a' content. (Metzinger 2003, p. 166)

In view of the facts that the practice of cognitive neuroscience relies on mutual constraints between higher and lower levels (see the next chapter) and that a strict vehicle-content distinction is the root of the symbol grounding problem (see 4.1) and insofar also of the weakness of classical theories of intentionality (see 6.2), we are now concerned with the vehicle-content distinction. Eventually it amounts to underpinning the negative claim that anything but a tight vehicle-content linking leads to too much problems.

The notion of a vehicle can be understood in two different ways: On the one hand, vehicles of content are the *bearers* of content. In a trivial and unspecified sense, that is to say that

vehicles of content are analogous to sentences that carry content. On the other hand, and distinguished from the first understanding, vehicles are not regarded as static bearers of content, but as processes. This notion of vehicle covers the subpersonal and attentionally unavailable neural architecture in which non-phenomenal and phenomenal mental processes are implemented, the globally available processes themselves, and both together. (Rowlands 2006, pp. 30-3)

If one thinks of the vehicle-content distinction as a relation between states within an information processing system (that is, the vehicles) and external objects or events (that is, the contents), then the distinction is faced with the same objections as the model of causal covariance (which is criticized later). Treating the content endowing objects as well as the relation between contents and vehicles itself as external to cognitive systems does not capture the intrinsic aboutness of our phenomenal episodes that are environmentally decoupled. Thus especially in the case of environmentally detached mental episodes, anything but a tight content-vehicle distinction is implausible, because the contents of those episodes cannot come from externally triggered input.

Could it not be that the model is adequate for non-phenomenal presentational and representational processing? Surely not, if one regards the non-phenomenally processed content in relation to the biological organism, and not as if the mental system processes organism-independent, “untouched” content (both in terms of physical stimuli belonging to third-person phenomenality¹⁵ and in the form of mind-independent objective stimuli). Akins (1996) showed that neurons do not per se respond to absolute properties of physical stimuli like temperature, but rather to changes in the actual state of the organism. For example, neural responding indicates that stimuli are warmer or colder than the dominant condition.¹⁶ Indicating such changes may provide precisely that information an organism needs to assess possible actions. It is clear that in this case neither the non-phenomenally nor the phenomenally processed content is organism- or mind-independent.¹⁷

It also interesting to ask about the motivation for externalising content. With a view to intersubjective objectivism and realism of cognition, it is convenient to externalise content. This seems to me to be the motivation behind it. As will be further articulated later, neither a strong intersubjective objectivism nor a strong realism of cognition is tenable when faced

¹⁵ Visual illusions show that it is not like that.

¹⁶ This idea can be comprehended by considering a simple example: The displeasing character of taking a cold shower fades over time.

¹⁷ In my view it would be wrong to think that neural responding that is not phenomenally experienced processes completely different information than phenomenally experienced responding. If this is right, then Akins' example a fortiori shows that the human organism does not respond to absolute properties of physical stimuli.

with empirical evidence. To be precise, avoiding a strong realism of cognition means both: humans have no access to mind-independent entities via phenomenal perception, nor is it the case that humans almost always classify phenomenal presentations and representations as most humans would do under optimal conditions, quite apart from the fact that being under those conditions is constitutive of any kind of mental content. In respect thereof it seems extremely questionable to view epistemological properties as in any way constitutive of phenomenal content.

I think preferring undistorted (optimal conditions) to distorted (resulting in misrepresentations) connections between vehicles and contents (Dretske 1983; Fodor 1990a) in order to adequately grasp the vehicle-content relation comes close to this. If one thinks of vehicles and contents as interrelated aspects of the same process (Metzinger 2003, p. 4), it is quite clear that in the case of misrepresentations or of representations of non-existent entities internal vehicles generate full-valued phenomenal content. Particularly phenomenal misrepresentations and simulations of non-existent entities should give rise to rethinking a strict distinction between vehicles and contents, the notion of cognitively impenetrable perception, and the reliance of higher-level mental occurrences on external stimuli. Moreover, even if one focuses on optimal conditions to state the vehicle-content relation, it is wrong to think of the phenomenal presentations and representations, to which conceptual simulations are applied, as containing mind-independent entities. However, it seems to me that binding to optimal conditions arises precisely from this naïve assumption. Insofar as one thinks of mental episodes under normal conditions as if they provided access to mind-independent entities, this takes on an epistemological touch. If one does not make such a naïve realist assumption, why it is so important then to identify optimal conditions with respect to defining mental content?¹⁸

In contrast to classical cognitive science accounts (for example, Fodor & Pylyshyn 1988) which start by comprehending vehicles that function in information processing systems and subsequently try to relate them causally to their external contents, neuroscientific accounts determine neural processes coming across as vehicles to sensorimotor systems that are already related to the environment (Bechtel 2008, p. 178). To be exact, both the neural responding to mind-independent entities (to be understood as those of physical theories) (to this extent neural vehicles generate a mental or phenomenal environment in the first place) and the neural responding to mind-dependent entities (e.g., in terms of concepts) (those that are visible to the naked eye) generates mental content. In respect thereof the sensorimotor

¹⁸ Given that one is not only anxious to ensure intersubjective objectivism.

vehicles are related to the environment.¹⁹ And even against the backdrop of such a phenomenal environment, by no means all of our parallel working mental occurrences causally co-vary with environmental objects.²⁰

If one deems only a tight interrelation of vehicles and content to be right, then the more specific question arises of how non-phenomenal or subpersonal processes generate phenomenal or personal episodes. The already mentioned synchronization of activation patterns makes a proposal in this respect.

Spivey (2007, p. 326) claims that a strict vehicle-content distinction implies the homunculus problem: Determining vehicles independently of the contents they are supposed to process raises the questions of *who* reads-out or processes the contents or especially in what way.²¹ Answering the latter question by deferring to further homunculi gives rise to an infinite regress. (cf. Damasio 1992) It also shows the explanatory weakness of referring to a central processor at all. Classical computational models are faced with this problem because they grasp cognition in terms of a central executive that processes language-like symbolic content. Actually, the virtual processor (vehicles) and the processed (contents) belong to the same collection of dynamic patterns. Vehicles and contents are two sides of the same coin – the ongoing dynamic brain-body-environment coupling generates more or less phenomenal and non-phenomenal content that is often intrinsically tied to neural, bodily, and environmental aspects. (Spivey 2007, pp. 324-8) More about this later.

By emphasizing the multilevel character of cognitive neuroscience, the following chapter will play an important part in contributing toward envisaging a tight connection between vehicles and contents.

3 The mechanistic framework

Higher level entities and activities are [...] essential to the intelligibility of those at lower levels, just as much as those at lower levels are essential for understanding those at higher levels. It is the integration of different levels into productive relations that renders the phenomenon intelligible and thereby explains it. (Machamer, Darden, & Craver 2000, p. 36)

¹⁹ Both ways of being related to the environment are believed to generate meaningfulness.

²⁰ It seems to me that principally all accounts which conceive of the contents as mind-independent and thereby strictly separate vehicles and contents are in need of correction. I think this is applicable to all accounts of causal co-variance (Dretske 1981, 1988; Fodor 1990) and of conceptual role semantics (Block 1986; Harman 1987). Apart from the reasons given so far, I will criticize such accounts in the second section of Chapter six.

²¹ I read Spivey's claim as suggesting that an externalisation of mental content raises the question how those external entities are intelligible within the frame of mental systems.

3.1 The metaphysical mind-body-environment problem

At the beginning of the first chapter we referred to a statement of Polger according to which the decisive question of the current mind-body problem is *how* the mind is part of the natural world. If we have a look at the accepted notion of supervenience, it is apparent that it gives no answer to this very question.

Since the supervenience relation itself is a purely formal notion, the subvening properties can range from neural to non-neural bodily to environmental properties. The supervenience relation is neutral on the issue of reducibility and the supervening properties asymmetrically covary with the subvening properties. (Walter 2007, pp. 135-8) The metaphysical mind-body or mind-body-environment problem is to explain why and how the mind supervenes on bodily or bodily and environmental structures.²² Simply claiming that the mind supervenes on the physical does not answer the metaphysical problem, it merely states the problem (Kim 1998, p. 13-4). Phenomena are only explanatorily reducible if the supervenience relation is not only ascertained, but if it is also explained why (Walter 2007, p. 147; Horgan 1993) and how the relation exists. The embodied mind paradigm and neurophenomenology precisely achieve the fleshing out of the supervenience relation. Describing mechanistic explanations provides further methodological comprehension of cognitive neuroscience.

3.2 Central properties of mechanistic explanations

The mechanistic framework provides an adequate description of neurobiological explanations. Furthermore, we will show that mechanistic explanations meet conditions that constitute good explanations.

In the following I will outline central properties of mechanistic explanations with an example from neurobiology, namely, the spatial memory of mice. Broadly speaking, spatial memory is the capability to learn to navigate through an unknown environment.

To describe a mechanism is to explain how a phenomenon is generated (e.g., Bechtel 2008; Bechtel & Abrahamsen 2005; Machamer et al. 2000; Darden 2006), how a task is accomplished (Bechtel & Richardson 1993) or how the mechanism as a whole behaves (Glennan 1996).

²² If the supervenience claim is limited to the mind-brain relation in the actual world (purely hypothetical), and if the supervenience claim is said to allow the multiple realizability of mental occurrences and leave room for non-identity, then one could reply in the following way: As we shall see later, to the extent that, if neural and phenomenal properties are specified at certain levels, identity claims are intelligible, and insofar as mental occurrences are not multiple realizable at all levels, this notion of supervenience is questionable.

What are mechanisms? Clusters of entities and activities organized in the generation of regular changes from set-up to termination conditions (Machamer et al. 2000, Chapter one). Examples of neurobiological entities are pyramidal cells, neurotransmitters, brain regions, and mice. What these entities engage in are the activities: pyramidal cells fire, neurotransmitters bind, brain regions process, and mice swim in water while searching for a means of escape. The entities and activities that compose the mechanism are organized, in a manner so as to do something or perform some function. This activity or behaviour of the mechanism as a whole is the phenomenon to be explained in terms of describing the mechanism. Examples of neurobiological phenomena are the mastery of a language, the storage of spatial memories, the release of neurotransmitters, or the generation of action potentials. A first aspect of this organization is temporal in nature. The stages of the mechanisms proceed in an orderly manner from the beginning to the end. Complete descriptions of mechanisms exhibit *productive continuity* – the course from set up to termination conditions is complete (or ‘gapless’). Productive continuities between the different levels make their explanatory connections intelligible. (Machamer, Darden, & Craver 2000, p. 3) Secondly, mechanisms are spatially organized. The components of the mechanisms are often compartmentalized, which allows us to individuate stages in a natural way. In our example, there are pre- and post-synaptic components of the mechanism of long-term potentiation (LTP). The thought of LTP is that the simultaneous activation of pre- and post-synaptic neurons strengthens their synapses. Moreover, components are localized, such as the hippocampus in the mechanism of spatial memory. If one has analysed the components of the hippocampus, one can describe connections between them. LTP is often studied in the mammalian hippocampus – an entity centrally involved in the mechanism of spatial memory. The hypothesis is that spatial memories are formed by changing the synaptic strengths between neurons in the hippocampus, and LTP is thus embedded into the mechanism of spatial memory. A third aspect of the organization of mechanisms is their hierarchical character. This makes our example pretty clear: Describing the mechanisms comprises mice learning to navigate, the hippocampus generating spatial maps, synapses inducing LTP, and macromolecules bending and binding. Those hierarchical descriptions require the integration of entities and activities at different levels. On the one hand, these components have to be attributed a functional role within the mechanism of the phenomenon to be explained (upward looking) (the question of the role of LTP in the mechanisms of spatial memory). On the other hand, the components have to be explicated in terms of lower-level mechanisms (downward looking). This can be continued as long as we obtain bottom-

out entities (in molecular neurobiology there are macromolecules, smaller molecules and ions) and their activities (geometrico-mechanical, electro-chemical, energetic, and electromagnetic). Crucially, each new decomposition exposes a lower-level mechanism until we obtain elements for which decomposition is no longer possible. If one does not find such a mechanism for the component, then there is a gap in the productive continuity of the mechanism. Hence, integrating multilevel mechanisms requires both the contextualisation of an item within higher-level mechanisms and the elucidation of that item in terms of lower-level entities and activities. Those *multilevel considerations* are characteristic of mechanistic explanations. (Darden et al. 2006, pp. 45-7)

I would now like to make an important point concerning evolutionary explanations. Evolutionary explanations narrow down to *why* explanations. If one says that intended offline simulation has evolved because it enables more complex forms of action planning, then *how* a human neurophysiology (blending out other explanatory paradigms in psychology) brings forth environmentally detached offline simulation remains open or in need of explanation.²³ To put it briefly, why explanations are not how explanations. (Cummins 2000, pp. 27-8)

What concept of causation do proponents of the mechanistic framework adhere to? They speak of *interlevel causation* (both bottom-up and top-down) *without interlevel causes*: To understand the thought, we have to make clear what it means to speak of *mechanistically mediated effects as hybrids of causal and constitutive relations*. The respective lower-level organised activation of entities simply constitutes the higher-level phenomena (e.g., the hippocampus generates spatial maps). Thus the *interlevel* relations take in the synchronous constitutive relations of the hybrid. On the other hand, the hybrid is complemented by etiological causal relations *within levels* (e.g., the binding of glutamate to the NMDA receptor). Top-down causation means that, for example, environmental stimuli causally affect the whole mechanism (this corresponds to intralevel causes) that is constituted of parts (this corresponds to interlevel causation). The change of the parts is a necessary condition for the change of the whole mechanism. In the case of bottom-up causation, changing a part in the mechanism causes the alterations of other parts in the mechanism (this corresponds to intralevel causes). This alters the mechanism as a whole (this corresponds to interlevel constitution) which might then cause changes in its environment (this corresponds to intralevel causes). Hence, mechanistically mediated effects are hybrids of constitutive interlevel relations and causal intralevel relations.²⁴

²³ As we shall later, Barsalou's PSS connected with Damasio's theory of convergence zones provide a how explanation of environmentally decoupled offline simulation.

²⁴ Correspondingly, some mechanistic explanations are constitutive, because they explain a phenomenon by describing its underlying mechanism, and some are etiological, because they explain an event by describing its

Mechanistically mediated effects, as hybrids of causal and constitutive relations, allow us to avoid well-known problems with causation (Craver & Bechtel 2007). Accordingly, causal relations are exclusively intralevel, whereby different levels are bridged by constitutive relations. Since causal relations occur within both higher and lower levels, lower-level accounts do not have to be viewed as primary and higher-level accounts as superfluous. Because interlevel relations are constitutive and synchronous, the higher-level phenomena do not have to be effects caused by temporally preceding and independent events. Constitutive interlevel relations also avoid the problem of *overdetermination* – changes in the behaviour of a component can only be caused by events internal to the mechanism, not additionally by factors affecting the mechanism as a whole. In other words, macroscopic stimuli can only affect the mechanism as a whole qua being constituted by the organisation of lower-level components. (Bechtel 2008, pp. 152-5)

What concept of level does the mechanistic framework adhere to? Determining levels by means of mechanisms begins with identifying the mechanism in terms of the phenomenon it brings about. At a next lower level, the working parts of the mechanism are assigned. Consequently, levels are only identified with reference to a certain mechanism – they do not spread over the natural world. Hence, this account of levels does not try to answer whether dogs, for example, are at a higher level than valleys, because they are not working parts of a common mechanism. This *local nature of levels* also appears at still lower levels, whereby the question of whether the sub-parts of two decomposed components (that were originally the working parts) are at the same or at a different level cannot be answered because there is no account of how they are combined in bringing about a common component. An important consequence of this local account of levels is that it does not aim at a causally closed and comprehensive lower-level, because lower-level components function as lower-level components only in relation to higher-level phenomena.²⁵

Moreover, Kim's causal exclusion argument against higher-level causes²⁶ is not true of levels of mechanisms²⁷ (1998, p. 84).²⁸ Why? The lower- (P) and higher-level (M) variables do *not compete* as sufficient causes and explanations. The behaviour of the mechanisms as a whole (M) does not compete with the organised behaviour of its components (P) as the cause of any

antecedent causes (e.g., dehydration belongs to the etiological explanation of thirst) (Craver 2009, pp. 107-8).

²⁵ Clearly, this proposal is distinguished from sorting out levels in terms of scientific disciplines. Moreover, the interacting working parts that build a level do not have to be of the same size (e.g., cell membranes as large parts can interact with individual sodium ions as small parts).

²⁶ Kim's argument is said to show that properties at higher levels of realisation have no causal properties over and above those of their realisers.

²⁷ Levels of mechanisms are understood as part-whole relations – the organisation of lower-level parts constitutes the whole mechanism.

²⁸ That is not to say that Kim intended to apply this argument to levels of mechanisms.

downstream processing. Indeed, the lower-level variables constitute the behaviour of the mechanisms as a whole. Moreover, the causally relevant processes of lower-level components are specified subject to higher-level phenomena. In respect thereof it does not make sense to say that non-organised realisers have all causal properties that the realisers at higher-levels are supposed to have.

Importantly, by virtue of their organisation, mechanisms can produce behaviour that their parts cannot do individually (that is, non-organised): The primate visual system, for example, responds differentially to motions, shapes, and objects. Generating patterns of activity which correspond to features of visual stimuli (from the viewpoint of third-person phenomenality) (e.g., a bouncing yellow tennis ball) requires the organisation of a host of neurons across several functionally distinct brain areas. Thus the organisation of components in a mechanism enables the mechanism to be influenced by environmental aspects which thus cannot impinge on the components individually.²⁹ In addition to that, there are explanatory generalizations comprising contrastive relations of causal relevance³⁰ that apply to the realised variables but not to the realizers: If, for example, a pigeon pecks because the paper is red, then changing the paper from scarlet to crimson does not change the pigeon's behaviour, though it does alter the activation vector in its nucleus rotundus. (Craver 2009, pp. 196-227)

How can hierarchical mechanisms be tested? By either intervening (stimulatory or inhibitory) at the lower level and detecting at the higher level (bottom-up) or by intervening at the higher-level and detecting at the lower level (top-down). An example of the former is the case study H.M. (reported by Scoville/Millner 1957): After portions of his temporal lobes, including the hippocampus, were removed, he lost the ability to remember recent facts, although he was still able to learn new skills. Zola-Morgan and Squire (1993) concluded therefore that the hippocampus is an essential component in the mechanisms which realize 'declarative' memory. An example of top-down intervention are fMRI studies that show, for example, specific activations of the motor strip (tongue, leg, hand) while reading action words (lick, kick, pick).³¹ But more of that later. (Bechtel forthcoming)

²⁹ One can add to this that it underpins our interactionist or embedded view of neural processing (cf. Northoff 2004). The stimuli with which the brain is supposed to interact can be regarded as belonging to our phenomenal world model. By the way, the stronger claim that properties like motion and shape are primary qualities is quite intelligible – against the backdrop of our phenomenal world model.

³⁰ Importantly, that is only to say that higher-level phenomena are causally relevant, not that they exercise novel causal powers in a metaphysical sense.

³¹ Besides, here mechanistic explanation is understood as an epistemic activity of scientists. This can be justified by considering that many scientific explanations turn out to be false. What scientists delineate are phenomenal representations of a mechanism, not the mechanism itself. A related question is to ask about the way in which scientists represent mechanisms. Hegarty (1992) introduces the notion of 'mental animation' which expresses well the dynamic visual simulation of the activity of the mechanism. In addition to visual simulations, linguistic representations and inferences can be engaged. (Bechtel 2008, pp. 17-22) I would even go

Now let us shift attention to the normative part of the mechanistic framework. There are implicit norms about what does and does not count as acceptable constitutive explanation in cognitive neuroscience. These norms shall come across as epitomizing the thought that good explanations in neuroscience show how phenomena are situated within the causal structure of the phenomenally accessible world. With respect to the implicitness of these norms we attempt to provide both an adequate description of neurobiological explanations and criteria being supposed to evaluate explanations. Here I do not argue for the claim that these norms are indeed implicit in the practice of neuroscience.³² I only want to show that the mechanistic framework has a concept of explanation that meets these conditions and that the Covering-Law model does not do justice to them. These norms are as follows: (1) Mere temporal sequences are not explanatory; (2) causes explain effects and not conversely; (3) causally independent effects of common causes do not explain one another; (4) causally irrelevant phenomena are not explanatory; (5) causes need not make their effects more probable to explain them (cf. Cummins 2000, pp. 1-6).

In what way does the concept of mechanistic explanation satisfy (1)-(5)? Consider the following argument: Causal relevance is analysed in terms of the manipulationist view (cf. Woodward 2003). The manipulationist view satisfies (1)-(5): (1') The explanandum cannot be explained by a temporally preceded event, because one cannot manipulate the explanandum by intervening to change the preceded event. (2') Similarly, effects do not explain their causes, because one cannot change the past by intervening in present states of affairs. The same principle applies to (3') and (4'): The variable whose alteration does not lead to a manipulation of the explanandum does not belong to the explanation. (5') If the explanatory relevant variable is manipulated, then the probability of the respective explanandum should change.³³ This can be linked with an account of constitutive explanatory relevance that is understood as sufficient condition for causal relevance. What is meant by constitutive explanatory relevance? Mutual manipulability is characteristic of constitutive explanatory relevance: The explanandum phenomenon can be changed by intervening to change a component (as exemplified by stimulation experiments), or the component can be manipulated by intervening to change the explanandum phenomenon (as exemplified by activation experiments).

so far as to say that the entities and activations of the mechanisms are ontic components, but that any ontic components we regard as external and causally efficient belong to our phenomenal model of reality.

³² Craver shows this using the example of neurotransmitter release (2009, pp. 22-8).

³³ This account of causal relevance remains silent in regard to the ultimate metaphysical nature of causation.

Because the following objections to the CL model are quite familiar to many, I make it short: Firstly, this model cannot distinguish laws of nature from non-explanatory generalizations (that is, it does not satisfy (1)-(3)).³⁴ Secondly, it provides no account of explanatory relevance.³⁵ Thirdly, one need not demonstrate that a phenomenon was to be expected in order to explain it.³⁶ (Craver 2009, pp. 21-106)

The following remarks on multiple realisability are based on the ideas of the mechanistic framework.

3.3 Thoughts on multiple realisability

Since we deal with neurobiological information all along, the following deliberations also go into the matter whether certain kinds of multiple realisation are or would be sufficient for the irrelevance of neurobiology for the understanding of mental occurrences. Moreover, because variable embodiment – that is, the claim that basic differences in neurobiological equipment are sufficient for different mental lives – is one consequence from PSS (see 4.3), and due to the fact that it is *prima facie* incompatible with multiple realisation, now I am going to develop a notion of multiple realisation that is compatible with this implication of PSS.

Now, I would like to answer the following questions: (1) Is multiple realisability an empirically justified claim? (2) Why did or does the thesis of multiple realisability look

³⁴ Cummins also argues for the thesis that the CL model does not apply to psychological explanations: Psychological laws are laws *in situ* that specify effects, not explanatory principles. Psychology's primary explananda are capacities. Effects and capacities in special kinds of systems are explained by the structure of those systems. The deductive-nomological model of explanation states the subsumption of antecedence conditions under a general law as the explanans that is deducible from the explanandum. Hence, the CL model of explanation does not apply to psychological explanation – that is, psychological explanation is not subsumption under laws.

Actually a psychological law is the explanandum, not the explanans. A psychological law is the explanandum because it specifies an effect (e.g., the McGurk-Effect: the perception of an acoustic speech signal (e.g., *da-da*) is influenced by simultaneous observation of lips movements (e.g., *ga-ga*)). Nobody can seriously think that the McGurk effect explains the data it subsumes, though it is a well confirmed regularity or law. Laws of psychology are laws *in situ* – laws that concern a special kind of system that is specifically constituted and organized. They specify regular behavioural patterns distinctive of a certain kind of mechanism, that is, effects. (Cummins 2000, pp. 1-15) But do not understand me wrong, I am only saying that subsumption under psychological laws in the fashion of the McGurk effect are not neuropsychological explanations. This does not preclude that the mechanistic framework could develop a new kind of neuropsychological law whose subsumptive force would come to be a match for a neuropsychological explanation.

The primary explananda of psychological explanations, however, are not effects (that is, psychological laws) but capacities, such as to understand a language, to grasp other minds, to see depth, to plan, to have self-consciousness, or to present philosophical theories in the form of written inferential relations between propositions.

In summary it can be stated that the decisive merit of Cummins is the insight that psychological laws in the depicted sense specify at best *what* the minds does, not *how* the mind does what it does. This well suits the quotation of Polger that is written down at the beginning of the first Chapter.

³⁵ For example, although there is a generalisation that if the Na^+ concentrations had been high, then the opening of the Ca^{2+} would be more probable, this opening is indeed explained by a rise in membrane voltage.

³⁶ For example, Ca^{2+} channels frequently open under conditions in which their opening is improbable.

plausible? (3) Does non-reductionism entail the truth of multiple realisability? (4) Assuming that multiple realisability is true, would this be sufficient for the falseness of brain-mind identity theories? (5) Would the empirical truth of multiple realisation be sufficient for the irrelevance of neurobiology for understanding psychological³⁷ phenomena? At first, however, I am going to give a metaphysical account of the realisation relation. Without such an account, none of these questions can be meaningfully answered. Moreover, I will show that this metaphysical account adequately apprehends the neurobiological practice.

Since the realisation relation depends on the kind of function the realizer is supposed to realize, and there are various kinds of functions, the concern is not a single realisation relation (Polger 2004, p. 113). What we need are specific 'realisation theories' (Poland 1994) for specific mental functions or capacities of the human organism.

In order to provide a working account of a specific realisation relation, I am now going to outline a specific concept of function. This notion of function belongs to the mechanistic framework: The function of X is what X is supposed to do in the mechanism in which X operates. The respective function of X is not its overall causal role, but a subset of its causes and effects which are relevant for certain explanatory purposes. Such an individuation of X's causal role shows how X is spatially, temporally, and actively integrated into a higher level mechanism. Additionally, functions are hierarchically realised insofar as higher-level functions are brought about by the performance of subfunctions which are in turn composed of further mechanisms with more specific functions, and so on. This division depicts a hierarchy of mechanisms, as previously mentioned.

Descriptions of mechanistic realisers are always selective – the components are individuated with reference to the mechanism's whole behaviour (which also has to be selected).

Mechanistic realisations differ from material, aggregative, and structural realisations in the following way: Unlike mechanistic realisations, structural ones only determine the interaction and organisation of structural properties, not activities. In the case of aggregative realisation, the lower-level properties are merely summed, not organisationally connected. Descriptions of material realisations simply determine spatially bounded entities as being composed of certain materials.

Does realisation have to be compositional? The account of emergence held by proponents of the mechanistic framework differs from mysterious varieties of emergence insofar as they say that the whole is not greater than the parts, plus their being organised. (Craver & Wilson 2006; Craver 2009, pp. 216-7; Van Gulick 1993; Wimsatt 1997)

³⁷ 'Psychological' is synonymous to 'mental'.

Because there are different realisation relations according to different notions of functions, the question of whether psychological functions are multiple realisable cannot be answered per se. It only makes sense to ask whether a certain function is multiple realisable or not.

The question of whether a certain psychological function is multiple realised also depends on the level at which the realisers occur. It is quite possibly that a certain psychological function is multiple realised at a lower level, whereas the same function is uniquely realised at the next higher level. (Funkhouser 2007, p. 18) Clearly, we are interested in querying whether psychological phenomena are multiple realisable by neural processes which function at the same level.

So what necessary conditions have to be satisfied so as to be justified in speaking of multiple realisation? The same psychological phenomenon has to be realised by two different types of realisers which function at the same level (Shapiro 2004).

Before I answer the questions, let us consider two more final remarks: The claim of multiple realisability can only be justified or rejected by means of empirical evidence.³⁸ Secondly, our considerations are limited to the question of whether organisms of the species *homo sapiens* occurring on earth have psychological properties that are multiple realisable. (Aizawa & Gillett 2007, pp. 2-3)

(1) At first, I would like to reconstruct the 'methodological argument' (cf. *ibid.*) against multiple realisation: (P1) If higher-level properties are multiple realised by lower-level properties, then there is no inter-theoretic constraint between the sciences studying lower-level properties and those studying higher-level properties (no constraint principle), and we would not find brain mapping studies that compare the brains of diverse animal species. (P2) There are inter-theoretic constraints (including mutual revision) between psychology or cognitive science and the neurosciences and interspecific comparisons. (C) Hence, psychological properties are not multiple realisable by neural properties. (Bechtel & Mundale 1999; Shapiro 2000, 2004) Actually the entire project of cognitive neuroscience (cf., e.g., Bechtel forthcoming) corroborates the second premise. No doubt the first premise of the argument is the contentious claim. I argue for the compatibility of multiple realisability and inter-theoretic constraint (that is, the falsehood of the no constraint principle) and hence the inconclusiveness of the argument: Consider a very brief depiction of a case study of human colour vision: Properties and relations of amino acids and 11-*cis*-retinal realise the light absorbing property of the human green photopigment and this can be realised by distinct amino acids having the same absorption spectrum. Properties and relations of photopigment

³⁸ Using conceivability claims does not count, because they are, as we shall see later, epistemically unreliable.

cells, water molecules, etc. realise a cone's property of transducing light. This property of cones is realised by G proteins with distinct activation properties. Human colour processing is realised again in the human retina by properties of the cones, amacrine cells, bipolar cells, horizontal cells, and retinal ganglion cells. Human colour processing is multiple realised by variations in the cone mosaic and by a varied number of distinct cone opsins. What this case study is supposed to show is that neuroscientific explanations are inherently multilevel in nature, that there are inter-theoretic constraints in practice, and that psychological properties are multiple realisable at different levels (by virtue of the transitivity of the realisation relation). (Aizawa draft, pp. 12-24)

Secondly, let us consider two classical arguments for the truth of multiple realisation offered by Block and Fodor (1972): (i) brain plasticity³⁹ and (ii) the possibility of artificial intelligence. The problem with the first argument is that it is not sufficient for the truth of multiple realisation. Only if the substituting brain areas brought forth the same phenomena via a different type of mechanical realisation, could brain plasticity be a case of multiple realisation.⁴⁰ As for the second argument – the contentious claim is not whether AI can behaviourally simulate the human mind, but whether artificial intelligences with sensorimotor systems distinctly different from ours can realise the same mental processes (Barsalou 1999, p. 639).

We are now going to deal with four classical 'conceptual' arguments: (i) argument from computation; (ii) argument from machine functionalism; (iii) argument from functional analysis functionalism; (iv) Putnam's likelihood argument (1967). (i) Each argument which derives the independence of the neural implementation of cognitive systems⁴¹ from defining cognitive systems in terms of computational systems, and from defining computational systems independently of neural implementation (cf. Eliasmith 2002), is subjected to objections to computational models. In the fifth Chapter I will raise such objections. (ii) The claim that cognitive mental states are machine-functional states is an empirical hypothesis.

³⁹ The plasticity of the brain is supposed to show the multiple realisability of psychological processes: Areas that were originally used solely to process information of one modality, were adapted to process another modality's information, because the originally responsible areas were damaged. For example, Melchner et al. (2000) have explored lesion-induced neural rewiring in the auditory cortex of young ferrets. By severing the normal linkage between inferior and superior colliculus in the midbrain and the medial geniculate nucleus (MGN) in the left hemisphere, a lesion was induced in a part of the auditory system upstream from the auditory cortex. The result was that, over time, the MGN received projections from the retina and in turn passed that information to auditory cortex. Therefore the rewired ferrets could “see with their auditory systems”.

⁴⁰ This reply is somewhat underdetermined concerning its content. Nevertheless, I think that it rightly points to the fact that the objection per se is not sufficient for the truth of multiple realisability. Conversely, calling attention to this insufficiency does not preclude that the phenomenon of brain plasticity can be sufficient for multiple realisability against the backdrop of certain assumptions.

⁴¹ We confine ourselves to the question whether these arguments provide an adequate description of episodes which we call 'cognitive'.

Turing machines are defined by input-output relations and by relations between mental states which are purely syntactical in nature. Since different physical systems can be described with the same machine table (identical causal relations), multiple realisability is a natural consequence of viewing cognitive mental states as Turing machine-functional states. (Putnam 1960, 1967) What is wrong with this argument? As we shall see later in more detail, defining mental operations purely in syntactical terms is not true of causally efficient mental processes. Moreover, if such a functional isomorphism were to provide an adequate simulation of mental processes, an explanation why the mind is the kind of thing that can be duplicated by such a functional isomorphism would still be required. It is no less plausible to view something as a hurricane that satisfies a machine-functional description of a hurricane (Block 1978; Searle 1980; Sober 1992).⁴² Moreover, it does not provide an adequate description – I think the mere fact that mental operations are not purely syntactically grasped from a phenomenal point of view (e.g., phenomenal experiences of imagery) suffices for a rejection. (iii) Attempts to derive multiple realisability by adopting a functional attitude to the mind (Fodor 1968) are not conclusive, as they in no way exclude that a certain concept of function can be uniquely realised by neurobiological mechanisms that are appropriately individuated. (iv) Putnam makes the following assumptions: pain is a mental state present in mammalian, reptilian, mollusc, and conceivable extraterrestrial brains (PA); plus two opposite claims: on the one hand, the mind is type-identical with the brain (TI), on the other hand, the mind is multiple realisable (MR). Hence $P(\text{MR} \mid \text{PA}) > P(\text{TI} \mid \text{PA})$. Yet again, $P(\text{MR} \mid \text{PA})$ is not per se greater than $P(\text{TI} \mid \text{PA})$ – it depends on the granulation in which mental and neural processes become individuated (we will presently see what this means), and on whether Shapiro's two conditions⁴³ are satisfied. (cf. Shapiro 2004 pp. 13-33)

(2) There are two interesting reasons why the thesis of multiple realisability looked or looks plausible. Firstly, philosophers often applied double standards – psychological phenomena were individuated very coarsely (e.g., a functionalist-behavioural individuation of pain in the form of pain behaviour), whereas brain states (which may be considered as philosophical fiction, since neuroscientists speak of complex activation patterns that bring forth phenomenal experiences (Bechtel & Mundale 1999, p. 177)) were often individuated very finely (humans and other organisms clearly have different brains). It is thus not surprising that psychological functions are multiple realisable by brain states. If both sides are coarsely individuated, then multiple realisability is not quite probable; the same is valid for

⁴² This broaches the subject that the fact that a certain behaviour is computable does not entail that the behaviour itself is a computation. This is a famous objection to computationalism.

⁴³ That is, the same mental phenomenon (first condition) is realised by different mechanical realisers (second condition).

corresponding fine-grained individuations of both sides. Secondly, the context or frame of reference for developing psychological and neuroscientific taxonomies was often neglected. That is, it is only appropriate to speak about sameness and difference with respect to some consideration or other – for example, both humans and dogs can certainly whimper, but their phenomenal experiences accompanying their whimpering is likely to differ. (ibid., pp. 201-3)

(3) To answer the question of whether non-reductionism implies multiple realisability, one needs a clear idea of what reduction is taken to mean. Since there are also etiological causal processes at different levels, thereby allowing different scientific levels, you have independence without multiple realisation. Because we adhere to a mechanistic notion of realisation and explanation, we have the following view of reductionism: High-level phenomena like mental capacities or observable behaviour must be known to be constitutively explainable in terms of lower-level mechanisms.⁴⁴ In other words, one has to get an idea of the explanandum. So these phenomena are epistemically and heuristically (cf. Bickle 1998; 2003) indispensable. Even if high-level phenomena are uniquely realised at a certain level, this does not imply that the high-level phenomena are epistemically irrelevant. If this epistemic indispensability is considered as non-reductionism, then non-reductionism does not entail multiple realisability.

Additionally, with respect to the fact that the specification of neural processes is dependent on higher-level phenomena (that is, lower-level components are determined in relation to higher-level phenomena), higher-level phenomena are epistemically relevant. If it is assumed that this downward specification of neural processes provides a neurobiological ontology, and if the debate on the ontology of mental processes bases on this ontology, then the ontology of mental processes is epistemically and heuristically dependent on higher-level phenomena. Thus the ‘ontologically mind-independent’ entities would be the respective components that constitute a higher-level phenomenon. The advantage of such an understanding of neurobiological ontology is that it makes multilevel relations intelligible. What is more compelling, as already mentioned, is that the whole mechanism exhibits emergent behaviour (e.g., the visual system responds to motion) and there are contrastive causal relations applicable to the realised but not to the non-organised realisers. To this extent one could say that the behaviour of a mechanism as a whole is not reducible to its non-organised realisers. However, that does not entail that this whole behaviour is not uniquely realised by a type of a complex neural activation pattern (to be understood as being organised) which is already specified in terms of relations to macroscopically observable

⁴⁴ Clearly, that is not a reduction in the sense of reducing one theory to another.

environmental stimuli (e.g., the function of the complex neural activation pattern is to respond to motion).

(4) Since neurobiologists specify brain processes at various levels, asking for the truth of brain-mind identity claims per se does not make sense. As already mentioned, this multilevel character also implicates that the question of whether mental phenomena (put the case that they are identically specified) are multiple realisable depends on which lower level is chosen. Hence, the fact that mental phenomena are multiple realised at many neurobiological levels does not entail that they are multiple realised at all those levels. This again implies that endorsing multiple realisation at some levels is compatible with brain-mind identity at another level (Aizawa draft, p. 9; 2009, p. 2).

What's more, one could construe identity claims as hypotheses that are supposed to guide subsequent research: It is assumed that mind and brain are identical in order to integrate and improve neurobiological and psychological theories. To this extent the identity claim is not a conclusion, but a premise of neurobiological research. This heuristic hypothesis is justified by its predictive (e.g., property verification or property generation tasks) and explanatory (in the sense of constitutive explanatory relevance: stimulation and activation experiments) success. In respect of these experiments the heuristic usage of the identity claim is common practice in cognitive neuroscience. Importantly, the heuristic identity theory is a type identity theory.⁴⁵ (cf. Bechtel & McCauley 1999)

(5) Now I will give a reconstruction of the 'argument for strong psychological autonomy': (P1) If higher-level properties are multiple realised by lower-level properties, then there is no inter-theoretic constraint between the sciences studying lower-level properties and those studying higher-level properties. (P2) Psychological properties are multiple realised by neural properties. (C) There is no inter-theoretic constraint between psychology or cognitive science and the neurosciences. (cf. Aizawa & Gillett, pp. 18-9) Again, the first premise is the contentious assumption. Since the case study of human colour vision showed that psychological properties are multiple realised at many levels, and that inter-level constraints are common practice in neuroscience, the first premise is wrong. Hence, multiple realisation per se does not entail the irrelevance of neurobiology for understanding psychological phenomena.

⁴⁵ Bechtel and McCauley emphasize (1999, p. 70) that the practice of cognitive neuroscience adheres to the principle of the indiscernibility of identicals, as the converse of Leibniz's law (that is, the identity of the indiscernibles).

As we have shown that multiple realizability per se is compatible with a certain concept of type identity, what about other arguments against identity theories in addition to multiple realizability?

3.4 Some objections to other classical arguments against identity theories

Because we adhere to a certain kind of identity theory, we are interested in challenging other classical arguments against identity theories. These include the argument from explanatory gap, Jackson's Mary argument, Kripke's argument for the contingent truth of brain-mind identity, and Davidson's anomalous monism.

(a) The following considerations are attempts to bridge the explanatory gap.

What is the explanatory gap, at least in the sense of Levine (1983) and Chalmers (1995, 1996, 1997)?

What is Levine's argument? Taking the deductive-nomological model of explanation for granted, the explanans causally explains because it necessitates the explanandum. Thus, if mind-brain identity is contingent, it is impossible to explain the mind in terms of the brain. If mind and brain were identical, then the identity would not leave an explanatory gap. According to Levine, only 'reductive' explanations can close the gap. Such explanations are functionalist – firstly, we determine our concepts by identifying their respective causal roles, and then we empirically ascertain the mechanisms that underlie these causal roles. But, so the argument, our psychological concepts of qualitative conscious states cannot be analysed in terms of their functional roles. Hence, an explanatory gap is left and therefore the mind-brain identity is wrong.

What is the objection? The more the structures of the phenomenal side and the structures on the side of the brain are adjusted to each other, the more intelligible identity assumptions will be. The more we can ascertain structure within the phenomenal real, the better chance we have of giving neurobiological explanations. Both sides have to have corresponding degrees of granularity. (Bechtel & Mundale 1999) Moreover, as we will presently see, it is simply wrong that qualitative conscious states cannot be analysed in functional terms.

According to Chalmers, the difficult problem of consciousness is to explain why mental occurrences have phenomenal properties. In other words, the challenge is to explain why it feels like anything at all. He claims that regardless of how precisely we determine the respective neural mechanisms, we do not come closer at an answer.⁴⁶

⁴⁶ It could be that both versions of the explanatory gap identify consciousness with the classic concept of a quale (CI Lewis 1929). We have a much more inclusive concept of consciousness that corresponds to the view of the

According to an interesting hypothesis that has already been mentioned several times, the central evolutionary function of consciousness is to make certain facts globally available for an organism. That is, consciousness allows us to attend to them, to think about them, and to react behaviourally to them in a flexible manner which more (in the case of behavioural control) or less (in the case of cognitive processing) automatically considers the overall context. The fact that a world appears to you, allows you to grasp that there is an outside reality and that you exist as well. Moreover, you can consciously experience emotions and thereby discover that you have goals and needs; you can experience yourself as a thinking being; you can discover that there are other agents with phenomenal experiences. Consciousness as a virtual organ also allows you to grasp the notions of truth and falsity, and the difference between transparent and opaque phenomenal occurrences. Additionally, your unified global model of a single world provides you with a single frame of reference that makes the distinction between actuality and possibility (simulations) possible. (Metzinger 2009, pp. 57-62) I think that this functionalist neurophenomenology provides an answer to the question of *why* biological organisms have phenomenal experiences (contrary to Chalmers' claim), and analyses phenomenality in functional terms (contrary to Levine's assertion).

Moreover, the multilevel nature of mechanistic explanations shows how phenomenal and non-phenomenal processes can be bridged in neurobiological terms (productive continuity). How can non-functionalist neurophenomenology⁴⁷ play a part in contributing towards bridging the explanatory gap? Let us consider a neurophenomenological pilot study (Lutz et al. 2002): There are two different phenomenal experiences (phenomenological invariants) (first-person data) which underlie specific neural and behavioral patterns (third-person data). In the first cluster (A), subjects reported being prepared for the stimulation, undergoing a feeling of continuity when the stimulus occurred and an impression of merger behind themselves and the phenomenal percept. In the second cluster (B), subjects reported being unprepared, diverted, and having a strong feeling of discontinuity in the stream of their mental occurrences when the stimulus was presented. During (A), a frontal phase-synchronous ensemble arose early between frontal electrodes and maintained throughout the trial. This group of trials exhibited a relatively short average reaction time (300 ms). The energy in the gamma band (30-70 Hz) increased during the preparation period. In (A), the energy in the gamma band was always higher in the anterior regions during the pre-stimulus

phenomenological tradition (there is no phenomenality without intentionality).

⁴⁷ 'Non-functional neurophenomenology' only means that it is unlike Metzinger's functional analysis of consciousness.

period compared with cluster (B), whereby the energy in the low band was lower. This indicates that paying attention during preparation is characterized by an increase in the fast rhythms accompanied by a decrease in slow rhythms. By contrast, in the cluster (B), when the stimulus occurred, a complex pattern of weak synchronisation and strong desynchronisation between frontal and posterior electrodes was ascertained. Moreover, the reaction time was significantly longer (600 ms). This discontinuity of brain dynamics was strongly correlated with the subjective impression of discontinuity. This study is believed to show that first-person data can be linked to stable phase-synchrony patterns (measured in EEG recordings), and that subjectively reported states of perception and preparation modulate both kinds of third-person data (reaction times and EEG recordings). The focus of the pilot study was the the dynamics of the interplay between experiential context of the subject leading up to the phenomenal perception and the phenomenal perceptual event itself. Whereby their interdependency was especially interesting: On the one hand, the way the antecedent and ongoing experiential context of the subject determines how the stimulus appears; on the other hand, this phenomenal perceptual object again reflects the flow of experience. This global process-stimulus structure and its temporal dynamics stand for the endogenous, self-organizing activity of the embodied brain that is understood as an autonomous dynamical system. What does this experiment say about the explanatory gap? By augmenting our view of both first- and third-person dimensions of consciousness, and by creating experiments in which they mutually constrain each other, neurophenomenology contributes to narrowing down the epistemological gap between subjective experience and neural processes in cognitive neuroscience.⁴⁸ Currently, neurophenomenologists do not assert that they have given explanatory bridges, but that they have provided a scientific research programme for advancing that task. (Lutz & Thompson 2003)

(b) What about Jackson's Mary argument?

Firstly, it is questionable whether Mary can really know nothing about sensations and their properties, if she knows everything there is to know about brain states, precisely because cognitive neuroscience works multilevel by its nature.⁴⁹ Secondly, the two premises of the argument invoke two different concepts of knowledge. The fact that Mary has concepts, whatever they may be (since conception is normally largely grounded in phenomenal vision),

⁴⁸ Moreover, I think that this neurophenomenological study already shows that it is not true that most that empirical evidence could ever establish is the 'correlation' between mental and neural states. What's more, if one considers mutual manipulability as characteristic of constitutive explanatory relevance in the context of cognitive neuroscience, then activation and stimulation experiments exhibit constitutive explanatory relevance. In respect thereof cognitive neuroscience provides explanations instead of 'mere correlations'.

⁴⁹ This point concerns conceptual or theoretical knowledge.

and the ability to undergo phenomenal visual experiences do not coincide. Thirdly, and this is the decisive point, what seems to be out of the question is that it is a non-sequitur to conclude explanatory irreducibility of phenomenal experiences from phenomenal differences of mental systems (conceptual knowledge in distinction from more embodied phenomenal experiences). And this is sufficient for rejecting this objection to our concept of type identity. (c) Now let us shift attention to Kripke's argument for the contingent truth of brain-mind identity.

Firstly, the claim that it is possible to be in pain without being in a certain neural state is epistemically unjustified, since it is determined by conceivability. And, as we will see later (cf. 6.9), conceivability is of no empirical theoretical relevance. Secondly, the apparent contingency of brain-mind identity may be explained by the fact that we do not know the identity conditions hitherto. It should be added that this concerns the knowledge of phenomenal experiences in the same way – we do not have an infallible capacity to specify our conscious life. Individuating phenomenal experiences is a skill that has to be learned. (Polger 2004, pp. 51-8). Thirdly, since identity claims rely on intelligibility and an adequate description of neural processes, the talk of 'C-fibres' or scattered brain states does not play a part in contributing to identity claims. Fourthly, as already shown, there are concrete proposals for explaining phenomenal experiences like pain (cf. Hardcastle 1999).

(d) According to Davidson's anomalous monism, every causally interacting mental event is identical to some physical event (monism, token-identity). There can be no strict laws on the basis of which any mental event-type can be explained, predicted, predict or explain (anomalous).

Because neurophenomenological or neuropsychological identity claims are only intelligible in terms of types (e.g., synchronized activation patterns), speaking of token-identity has no purpose. As we said above, scattered tokens of brain states do not exist. That is not to say that the type-token distinction makes no sense, but that the notion of token identity is hardly intelligible against the background of current cognitive neuroscience. Moreover, because the subsumption under psychophysical laws provides no explanations (with respect to Cummins' concept of psychological law) and because the practice of cognitive neuroscience does not search for laws as principles of universal applicability, but for specific phenomena or behaviour brought about by specific mechanisms consisting of an organised interplay of components, the possibility that there are no strict psychophysical laws is irrelevant for the question of whether brain-mind identity is true.

While this chapter also was concerned with interlevel constraints on mechanistic explanations, Barsalou's theory of perceptual symbol systems (PSS) endorsed in the following chapter attempts to explain higher-level cognition in terms of sensorimotor processing. This theory may be considered as an instance of mechanistic explanations.

4 Barsalou's theory of perceptual symbols systems (PSS)

In an embodied mind, it is conceivable that the same neural system engaged in *perception* (or in bodily movement) plays a central role in *conception*. [...] Indeed, in recent neural modeling research, models of perceptual mechanisms and motor schemas can actually do *conceptual* work in language learning and reasoning. (Lakoff & Johnson 1999, pp. 37-8)

4.1 The symbol grounding problem and the Chinese Room

Because it is one of the main problems for theories of amodal symbol systems, now we take a look at the symbol grounding problem.

An usual way of obtaining the meaning of a word one does not know is to look it up in a dictionary. Yet it might be that one also does not know any of the paraphrasing words, nor any of their circumscribing definitions, and so on. This possible infinite regress, that can only be avoided if some words are grounded by resources other than ungrounded paraphrases, is known as the 'symbol grounding problem' (e.g., Harnad 1990, 2003). Looking for meaning in such a way is analogous to looking for meaning in a Chinese/Chinese dictionary when one does not have any knowledge of Chinese besides syntactical information (Searle 1980). Solving the symbol grounding problem requires endowing symbols with meaning without going back to an external interpretation or external entities. That is, the symbols have to be grounded within the cognitive system itself, if they are supposed to explain what we do when we think. (e.g., Harnad 1992, 1993)

What if one grounds mental content in sensorimotor simulation? Then, as we shall see later in detail, either the symbol grounding problem does not arise at all (cf. 4.3) or it is solved by linking arbitrary linguistic tokens to sensorimotor simulations (cf. 4.5). In the case of non-linguistic conceptual processing, the symbol grounding simply does not arise, because the sensorimotor simulations are intrinsically meaningful (cf. Cummins 1996, Chapter nine). Language plays a role in conceptual processing only to the extent that areas responsible for processing word forms (perception of language and speaking) are connected with content bearing sensorimotor simulations. Hence, because the needed contents come into play within

the cognitive systems themselves, be it non-phenomenally or phenomenally, the symbol grounding problem is overcome.

As phenomenal representations are intrinsically meaningful⁵⁰, and because offline simulations are activations of structures being extracted from phenomenal representations, sensorimotor simulations neither have to be interpreted by inner homunculi (homuncular fallacy) nor present themselves to an understanding system *as* having a certain content (communicative fallacy). Committing the homuncular fallacy is to assume that inner homunculi literally interpret such-and-such internal representations as standing-ins for such-and-such states of affairs. Indeed, there is no interpreting, but the causally efficient processing or operating is content-sensitive itself. Anyone who proceeds on the assumption that “a state [...] presents itself to an understanding system *as* having a certain content” (Wheeler & Clark 1999, p. 126) commits the communicative fallacy.

Damasio's hypothesis of convergence zones is introduced next, because it is required for the subsequent reconstruction of PSS.

4.2 Feature maps and convergence zones (Damasio's hypothesis)

The pivotal idea can be explained by the following function of convergence zones: Once an active pattern in a feature map is captured by conjunctive neurons in a CZ, these conjunctive neurons can later reactivate the pattern in the absence of environmental input (Simmons & Barsalou 2003, p. 455). Conjunctive neurons of CZs can only function as stand-alone representations, if they feed forward to automatically to linguistic responses under very routinised conditions, such as word associations. Non-automatised conceptual processing requires the activation of feature maps. (ibid., p. 456)

Each of the six sensorimotor modalities contains the same configuration of four *subsystems*: *Feature maps*, *analytic convergences zones (CZs)*, *holistic CZ*, and *modality CZs*. Every modality (olfactory, gustatory, motor, somatosensory, auditory and visual processing areas) includes feature maps that code the content of modality-specific states, e.g. colour is coded in visual processing areas, whereas somatosensory processing areas coding surface conditions.

In contrast to feature maps that are to a great extent independent of attention, analytic properties result from selective attention towards phenomenal perception. Furthermore, while feature maps provide the construction of perceptions, analytic properties support the representation of cognitive-level categories. Once the subject has turned its attention to a

⁵⁰ If one assumes that meaningfulness is not intrinsic to phenomenal representations, where does the meaningfulness of higher-level cognition come from then?

specific configuration of features in a single subregion (e.g., a doorknob), the features become tied together in an analytic CZ. Subsequently, these linked features build an analytic property which can be utilized for a multitude of conceptual tasks (e.g., inferences, categorization). The conjunctive neurons within analytic CZs are organized in accordance with the *similarity in topography (SIT) principle*, as a consequence of which the spatial proximity of two neurons in a CZ mirrors the similarity of the features they compound. Instead of organizing the analytic properties by category, the SIT principle arranges them by *property type* (e.g., shape, colour, movement). For instance, although crocodiles and elephants are both animals (at the superordinate level), they vary widely in respect of visual properties – following the SIT principle, these highly diverging property types should be broadly dispersed in the analytic CZs in vision. Conversely, if different categories share more characteristics of respective property types than instantiated properties within a category, then these inter-categorical properties should be located spatially more adjacent than intra-categorical properties. Given the case that a category has a relatively unique visual property type, then of course we have an instance of category-specific topographic organization inside analytic CZs.

Corresponding to analytic CZs, holistic CZs comprise conjunctive neurons that capture activation patterns in feature maps for holistic properties. Holistic CZs capture the configuration between manifold analytic properties, for example, the arrangement of eyes, nose, ears and mouth in a face. In the manner of the SIT principle, the topographical proximity of the conjunctive neurons reflects the similarity between the holistic properties they are encoding.

As distinguished from feature maps and analytic and holistic CZs, modality CZs neither process features, nor do they combine features to conceptual properties. Instead they capture correlations between diverse analytic and holistic properties. Although these property correlations are decisive for category structure, modality CZs configure category representations. Consider an example: Because different bottles have similar analytic and holistic properties, spatially close conjunctive neurons in modality CZs will capture them. Assume that this is respectively true of cans and dogs. Now as dogs vary more from bottles and cans than bottles and cans differ from each other, the conjunctive neurons for bottles and cans are relatively removed from those of dogs. Consider another example: Since the diversity of visual properties of artefacts is much greater than the one of animals, the conjunctive neurons that capture the property cluster of artefacts should be more dispersed.

Beyond the configurational similarity between the six modalities via sharing the same four subsystems, the cross-modal CZs connect properties between modalities. Cross-modal CZs do not rely on uniting the representational format because their conjunctive neurons simply detect the simultaneous firings of conjunctive neurons in the modality CZs. What is more, they are determined by the same principles as modality CZs, unless the conjunctive neurons of the cross-modal CZs couple the modality CZs, which again link conjunctive neurons in analytic and holistic CZs. Let us illustrate the relation between modality and cross-modal CZs by an example: Although real cars and toy cars have adjacent conjunctive neurons in visual modality CZs because they bear similar visual properties, their cross-modal CZs lie more remote, because real and toy cars differ significantly in somatosensory, olfactory and auditory properties. (AR Damasio 1989, AR Damasio & Damasio 1994, cf. Simmons & Barsalou 2003)

4.3 Reconstructing PSS

The pivotal idea of PSS is that feature maps of sensory-motor brain areas become re-enacted during offline cognition. Offline cognition is thus embodied. In principle it is no different than the already introduced notion of neural exploitation (see 1.3).⁵¹ PSS provides a characterisation of central higher-level cognitive phenomena, from intentionality through to concepts, understanding, thought, and language. As we will presently see, in contrast to McDowell (1994), the phenomenal world not only rationally influences our thoughts if our experience is already conceptually structured, but also if our thoughts themselves are re-enactments or simulations of perceptual experiences. If one objects that this idea confuses the subpersonal-personal distinction⁵², then it can be answered that the processes at the subpersonal and personal level can only be adequately understood, if they are specified interdependently – just what the mechanistic framework has in mind.

⁵¹ PSS's model of conceptual knowledge can be localised as follows: According to Barsalou, a single multimodal representation system of the brain (that is, feature maps) sustains various types of simulations across different cognitive processes (such as high-level perception, implicit memory, working memory, long-term memory, and conceptual knowledge) which are constituted by specific mechanisms. In high-level perception and implicit memory, convergence zones capture representations and subsequently initiate simulations that accomplish perceptual completion or repetition priming, for example. Conceptual knowledge utilizes the same representational system but controls it with the help of convergence zones in the temporal, parietal and frontal lobes. The working mechanism in turn employs the same representation system, but controls it by means of medial temporal systems and different frontal areas. Corresponding to PSS, simulation is the allying concept for diverse neural processes that control a single representation system. (Barsalou 2008)

⁵² That is to say that PSS is a theory at the subpersonal level, whereas McDowell's position is a theory at the personal level.

In the following, I will reconstruct the substantial components of PSS. These remarks are borrowed mainly from Barsalou's "Perceptual Symbol Systems" (1999).

Modal perceptual symbols⁵³ are analogically and non-arbitrarily linked to phenomenal perceptual states. What does this mean? They are modal because the same neural systems that underlie perception also underlie perceptual symbols, in other words, perception and conception have a common representational⁵⁴ system. Due to their modal character, they are analogical – to a certain extent phenomenal perceptual states and perceptual symbols have a common structure. Non-arbitrary linking to perceptual states means that the representational vehicles of perception and conception are isomorphic. By contrast, to the extent that words bear no systematic similarities (representational schemes and constitutive principles) to the phenomenal perceptual states they refer to, they are amodal. For example, whereas within amodal accounts symbols for *tables* have no similarity to tables, perceptual symbols for tables are similar to phenomenal perceptions of tables. Because the symbols in a perceptual symbol system are modal, they are non-arbitrarily linked to their phenomenal perceptual forerunners.

I will now outline five central properties of PSS. This is not aimed at specifying these properties in terms of concrete neural mechanisms, rather it should be considered as a '*high-level functional account*' (Barsalou 1999, p. 6; my italics) of how the brain could implement a conceptual system via sensorimotor areas.

Firstly, a perceptual symbol is a *record* of the neural activation that underlies perception. What is the difference between perception, conceptual representation, and imagery? Whereas imagery and perceptual symbols come quite close together according to PSS, and whereas the sensory-motor representations are more conscious and detailed in imagery than in conceptual processing (if one conceives of imagery as relatively conscious), perception fundamentally differs from processes engaging concepts. On the neural level, the causal influence of CZs to re-enact sensory-motor areas only amounts to conceptual processing, it exceeds perception; more about this later (see 4.9). Another important issue concerns the role of consciousness. Basically, perceptual symbols are unconsciously processable neural representations. Perceptual symbols also serve as material for more or less conscious mental

⁵³ To speak of 'symbols' is slightly misleading because one could be inclined to consider them as pictorial signs that have to be interpreted by a homunculus in order to be contentful. In fact they are intrinsically meaningful modality-specific bodily stand-ins that can be simulated.

⁵⁴ In this section I use 'representation' (or 'representational') to mean internal neural stand-ins. If they are activated in full-blooded conceptual tasks, they are supposed to correspond to multimodal feature maps. If these neural stand-ins are not activated, they are supposed to correspond to CZs. It would be wrong to think of non-activated features in terms of CZs as non-activated full-blooded representations – CZs merely capture the activation of feature maps in order to be able to re-enact multimodal bodily states in the absence of environmental input.

simulations. Research on skill acquisition shows that consciousness falls away in the course of routinization.

Secondly, perceptual symbols are *schematic*. The record of the neural activation realizing the phenomenal perceptual state does not grasp the total brain activation of the phenomenal percept, but rather a small subset of neurons represent a part of it. The symbol formation process explains the schematic nature of perceptual symbols: A subset of neurons that underlie phenomenal perceptual states is *selected* and *stored* in long-term memory. Quite possibly, conscious experience may be necessary for the first occurrence of this process. How do perceptual symbols function once they are stored? Since perceptual symbols are associative patterns of neurons, they exhibit a *dynamical, non-discrete* character: If other perceptual symbols are subsequently stored in the same CZs, the connections in the original pattern will be altered, whereby possible future activations would be changed. Additionally, because different contexts can bias activations towards some particular features in patterns, the original pattern could crucially be changed. In terms of dynamical systems theory, a perceptual symbol is an attractor state. As the network changes (connections), the attractor changes. Context and activation of attractor co-vary. To close the circle, as the attractor changes, the network changes.

Need the construction of the schematic representation of shape be integrated into a holistic image, comprising orientation, colour, shape, and so on? Not at all. Inasmuch as perceptual symbols are defined as unconscious neural representations, the perceptual symbol for a particular shape could stand in for the shape *componentially*, while perceptual symbols for other dimensions remain inactive. Even if phenomenal experiences accompany perceptual symbols, they do not have to be holistic – in terms of being irreducible to schematic components.

Furthermore, in virtue of schematicity, it would not be surprising if a phenomenal individual is never simulated completely. Moreover, the forgetfulness and reconstructive vein of human memory do not allow accurate remembering.

Thirdly, perceptual symbols are *multimodal* in nature. The symbol formation process operates not only on vision but also on the other four modalities – audition, haptics, olfaction, and gustation – and on proprioception and emotion. Just like subjects acquire perceptual symbols for speech from audition, they also acquire symbols for colours from vision, symbols for taste from gustation, or symbols for leg positions from proprioception (and vision). Fourthly, related perceptual symbols become organized into a *simulator* that allows the construction of a potentially infinite set of *simulations*. What does this mean? Consider

the symbol formation process while viewing a particular racing bicycle. As one regards the racing bicycle as a whole, forming symbols of the chainstays, the handlebars, the fork, and the tyres, neural records become selected and integrated into the spatially organized system. As one subsequently looks at the handlebars, further individual symbols for the handlebar tape, the stem, and the brake levers become integrated into the spatially organized system. This spatially related coding enables coherent simulations during offline processing. Additionally, imagine a specific event sequence, for example, riding the bicycle. Imagine that you grasp the handlebars, climb on the saddle, and finally begin to pedal. Due to the fact that the symbols for each subevent are stored in a temporally related way, one can later simulate this event sequence. This event simulation will probably be multimodal, including, for instance, the proprioceptive information on pedalling or sitting on the saddle, the auditory information of the rattling chain, and the haptic experience of the blowing wind. Importantly, the symbols extracted from a phenomenal entity or event become integrated into a *frame* (using the example, a frame of a racing bicycle) that comprises a great amount of multimodal content. A frame is seldom if ever experienced entirely, depending on which context subsets of the frame become active so as to simulate a specific experience. In conclusion, a simulator consists of a frame that integrates perceptual symbols across a variety of category instances and the dispositional property of generating endless sets of simulations.⁵⁵

According to PSS, a simulator is a *concept*, the knowledge and the attended processes that allow subjects to adequately represent types of phenomenal entities or events. An adequate representation for a table, for instance, means in general that the subject can simulate multimodal experiences which are predominantly shared in a culture. The different simulations, retrievable by a simulator, are viewed as different *conceptualizations*. Meaningful components within our phenomenal representational system involve individuals and categories (e.g., natural kinds or artefacts). A *category* is a set of individuals belonging to our phenomenal environment or to introspection. A perceived entity belongs to a category if a simulator for that category can produce a satisfactory simulation of the phenomenally perceived entity, which is called successful *categorization*.

What about *concept stability*, intra- and intersubjectively? Is it possible at all, considering the widely diverging simulations of categories within and between individuals, prevalently influenced by context (including environment and currently active contents of the respective individual)? Intrasubjectively, if one simulator of a category produces different simulations,

⁵⁵ Running simulators is not only bound to the empirical course of sense impressions, the processing of space, objects, movement, and emotion is also genetically predisposed. Naturally, the actualization of these dispositions also substantially relies on interactions with the environment.

they can be viewed as exemplifying the same concept. If different subjects have similar simulators, then we also need not abandon concept stability between subjects. Are there reasons for this? On the one hand, contextual constraints during communication lead to similar simulations between subjects. Furthermore, a similar neurobiological equipment, a shared phenomenal physical environment, and conventions brought about by socio-cultural institutions contribute to stability. It is very possible that subjects have the principal possibility to simulate the other person's conceptualization, adequate constraint allows proper intersubjective coverage. Clearly, to determine whether intersubjective stability is the case or not depends on the level of granularity – in the strict sense, nobody shares concepts with someone else, because every subject made specific experiences leading to specific context-sensitive re-enacted concepts.⁵⁶ Let us consider an experiment that examines the degree of consensus between pairs of individuals for typicality judgements of different category members (e.g., different typical instances of [BIRDS]): Across a variety of studies, the degree of consensus was only about 40%. Intrasubjectively, when the same subject judged the typicality of the same instances twice, at an interval of two weeks, the corresponding average was around 80% (as reviewed in Barsalou 1993). With respect thereof, the representations of categories vary both inter- and intrasubjectively, depending on the situations the people are anticipating, which again depends on the whole context in which the people are embedded at a certain time.

The concepts can also be similar because the experienced entities allow similar affordances (e.g., many people share the concept of a CHAIR insofar as it contains a simulation of taking a seat).

Fifthly, as already mentioned, a *frame* integrates perceptual symbols into a spatially organized system in order to create specific simulations of a category.

Let us now look at four core properties of frames: attributes, predicate-value binding, constraints, and recursion. Predicates are more or less identical to unspecified frames, considering our example, the comprehensive representation of the racing bicycle depicts a predicate. Predicate-value binding occurs if different specializations (e.g., handlebars for time trial or normal handlebars) become tied to the same subregion (handlebars), hence, different values become bound to the same predicate. The associated activity between different specializations (e.g., handlebars for time trials and smaller front wheel), thereby simulating specific bicycles, signifies the property constraint. In virtue of simulating within a specific simulator, recursion arises. Initially one only simulates the schematic overall shape

⁵⁶ This indirectly shows a great potential of an multimodally embodied notion of the mind: It allows a quite fine-grained specification of non-phenomenal and phenomenal mental episodes.

of the handlebars, subsequently one focuses on more specific components, such as the brake levers or the handlebar tape.

If one considers conceptual relations (at least in respect to the ordinary psychological representation of such related contents) as a case of background knowledge, PSS has two ways in which background knowledge specifies the content of concepts, namely, framing and background meaning. In framing, specifying the content of a focal concept necessarily depends on the content of another concept, for example, *hand* is specified relative to *arm* and *body* or *handlebar* is specified in relation to *bicycle*. In background-dependent meaning, the content of the focal concept co-varies with meaning of the background concept. Changing the background from *human* to *bear*, the content of the focal concept *arm* changes.

The following remarks deal with additional properties of PSS that can be derived from the core features so far presented here: productivity, propositions, variable embodiment, and abstract concepts.

How then can a perceptual symbol system be *productive*, in other words, how can it produce a principally infinite amount of representations from a finite number of symbols and thereby transcend experience? Because built schematic representations are enriched with once filtered out types of information – for example, the simulator of a racing bicycle, lacking colour and labels, produces a comparatively rich bicycle simulation with the help of a simulator for blue and some for labels – the process of productivity could be understood as the reverse of the symbol formation process (schematization and specialization are complementary). However, productivity does not amount to nothing more than filling-in already established schematic regions, it can also arise from substitutions, transformations, or deletions of an encountered structure. It is possible to simulate never seen entities, such as a flying table via a simulator for an event sequence of flying and one for a table. Are there any constraints on the production of simulations? There are some, which can be traced back to certain affordances. Consider an example: It seems almost impossible to link the perceptual symbol for *running* to a simulated football, because the football lacks a critical feature, namely legs. Furthermore, the spatio-temporal features of the simulation itself generate emergent properties – for example, imagine the specific way a simulation of a running table might look. Language also contributes to make productivity possible: If a friend tells you of an exciting experience you did not undergo, you can understand his story only in virtue of

linguistically-initiated conceptual simulations.⁵⁷ Given that you do not need to experience everything first hand in order to acquire information, PSS exhibits productivity.⁵⁸

Let us now shift attention to the capacity of PSS to implement *propositions*. Firstly, consider type-token mappings. Imagine you see a bottle standing on a table. If you bind a simulator for a bottle, for instance, via a simulation of a bottle with the phenomenally perceived individual bottle, then you have all you need for instantiating a type by a token. Naturally, the mapping can be true or false – to simulate a cup and bind it with the phenomenally perceived bottle would be a false mapping. Since perception and simulation reside in the same representational format, the final phenomenal representation of the bottle is a fusion of both, bottom-up and top-down information. Secondly, mapping a simulator onto a phenomenally perceived individual allows one to draw categorical inferences. If the bottleneck is occluded, but the cap and the rest of the bottle visible, the binding of your bottle simulator enables you to infer that there is a bottle. Or imagine perceiving the bottle attached to a rope that steadily moves the bottle up and down. If you turn away for a moment, and then simulate the trajectory of the bottle, you can predict its position when it reappears. A given phenomenal perception can be manifoldly interpreted, just as you could map your simulator for a table onto the perception.⁵⁹ A hierarchical simulation of the bottle upon the table could be constructed productively from the simulators for *bottle*, *table*, and *upon*. With reference to my first Chapter (see 1.2), the simulator for *upon* may be considered as a good example of an image schemata. Another way in which simulators yield different interpretations of the same phenomenal perception consists in shifting focus from one part of the hierarchy to another, in the example, using the simulator for *below* to produce a simulation of the table below the bottle. Typically, propositions capture a gist that can be paraphrased manifoldly – in PSS captured by the fact that different sentences (“a upon b” or “b below a”) can initiate the same simulation.⁶⁰

Next we will deal with the opposite ‘SETI’ version of multiple realisability⁶¹, namely, *variable embodiment*. This implies that the mind varies with the biological system that brings them forth. To put it another way: Changing the biological system is sufficient for changing the system’s mental processing. Variable embodiment is a natural consequence of PSS.

⁵⁷ According to Langacker (1997), the productive combination of adjectives, nouns, and other linguistic components corresponds to the productive combination of properties, individuals, events, and other conceptual elements. This coincident productivity is a necessary condition for the human ability to understand communicated sentences in the absence of the respective entities.

⁵⁸ There is a pertinent illustration of how PSS implements productivity in the appendix (upper illustration).

⁵⁹ That is not to say that the phenomenal representation is not meaningful unless the simulator is mapped onto it.

⁶⁰ There is a pertinent illustration of type-token mapping in the appendix (lower illustration).

⁶¹ “SETI MRT: Some creatures that are significantly different from us in their physical composition can have minds like ours.” (Shapiro 2004, p. 7) “SETI” stands for the Search of Extraterrestrial Intelligence project.

How? Since the way in which the brain creates phenomenal perceptions has a direct bearing on simulational content, because symbols and percepts rest upon the same representational format, one can speak of variable embodiment. As we have already seen (see 3.3), such a strong neural constraint still allows certain kinds of multiple realisation.

Variable embodiment fulfils at least two important functions: adapting simulations to specific environments, and warranting that different individuals optimally map their specific simulations to phenomenal perception.⁶²

Finally, let us look at PSS and *abstract concepts*. Irreconcilable opposites? Not at all. We concentrate on analysing the way of representing the everyday sense of TRUE, exemplarily. How is truth represented? Imagine a speaker says, “Look, there’s a bottle bobbing up and down above the table behind you.” Almost automatically, you construct a simulation of the event sequence described. Subsequently, you turn around and try to map your simulation to the phenomenal perception of the event. You then judge whether your simulation adequately represents the perceived situation. If it is the case, the speaker’s sentence is true. What is the simulator for truth then? After repeatedly performing the event structure of simulating, focusing, and comparing simulation and phenomenal perception, individuals learn to simulate the experience of successful mapping, and such simulations could constitute people’s concept of truth.

Barsalou and Wiemer-Hastings (2005) give evidence for the following four hypotheses: (1) Concrete and abstract concepts are reliant on situational content. Just as you have to know the application, the usage, and so on of hammers to understand the concrete concept HAMMER, so a complex situation is also necessary to represent TRUE. (2) Concrete and abstract concepts differ in situational focus. Concrete concepts are processed by focusing on objects against their background situations. Abstract concepts, however, are aimed at events and phenomenally opaque properties. (3) Abstract concepts are more complex than concrete concepts. Concrete concepts are quite localized, confining themselves to spatially limited situations (hammers occupy a certain region). The contents of abstract concepts, by contrast, are distributed across situations. Moreover, their being extraordinarily related to

⁶² Imagine that two individuals interact from time to time with specific tokens of the same type of coke can in different contexts; the one uses it as container for watering flowers, whereas the other one takes it to decorate his living room. In perceiving their respective cans, they have more or less differing perceptual symbols to represent them, depending on their respective behavioural functions. What about the second function? Exemplified by colour vision, due to the fact that different subjects have slightly differing colour perceptions, and that subjects’ simulators for colours are grounded in finely discerning perceptions, their symbols also subtly differ. However, precisely this fine intrasubjective tuning at best allows matching symbols to perception during categorization. Such variation should also arise in phenotypic characteristics other than colour, such as shape, smell, or movement.

phenomenally opaque states may play a part in contributing to this complexity. (4) The content of abstract concepts can be simulated. Event structures as well as phenomenally opaque mental episodes (e.g., emotions) can be readily simulated in multimodal terms.

Is the claim that abstract concepts are directly experienceable not incompatible with what was said before, namely, that abstract concepts are metaphorically grounded in sensorimotor domains (basic-level and image schematic structures)? No. One could say that metaphors only augment the contents of abstract concepts (Boroditsky & Ramscar 2002), whereby direct experience of abstract concepts is nevertheless central to their content.

The following section aims at connecting dynamical systems theory and PSS.

4.4 PSS and dynamical systems theory: context sensitivity, embodiment, and temporality

It often happens that the same type of input – conceived of as a focal point in the environment – is processed in different ways, because active inner processing and peripheral environmental characteristics form a different context. This dynamic variability is noticeable, neurally as well as phenomenally. In representational and simulational terms, on a given occasion those features of a category become active that have been processed most frequently and most recently and that are to this extent most closely associated with the current context. These activated features make up only a small subset of the variable contents that a person has gathered over a lifetime. Even if simulators are stable in a sense, features can be retrieved, combined, and applied in such ways that they nicely suit the dynamicist's intuition that minds are never in the same state or process twice.

What about embodiment? For dynamicists, brain-body-environment interactions are central to almost all intelligent behaviour. By contrast, disembodied representations are language-like codes involving nothing that suggests their grounding in perception, body, and environment.

PSS is embodied, since conception and perception share a common representational system. PSS is inherently representational insofar as conceptual simulations are mapped onto phenomenal perceptual representations, and insofar as permanently changing stand-ins for aspects of phenomenal representations are used in offline simulations.

Since simulating a category consists of having multimodal sensorimotor knowledge (e.g., of a racing bicycle) that enables the generation of specific simulations of that category in specific situations, context sensitivity is a natural consequence of simulation. Context-

sensitive retrieving of relevant simulations may be a question of statistical frequency and resemblance between simulations and phenomenal representations. Depending on the respective context, a specific simulation is tailored to it. For example, representing racing bicycles that were used in the 1980s looks differently to representing ones that are used today. Similarly, representing racing bicycles being ridden may generate simulations from the cyclist's perspective, while representing racing bicycles being repaired may create simulations of a racing bicycle's shifting system or its brakes. Insofar as PSS's simulations evade the costly transduction into amodal symbols, they are at least more capable of satisfying real-time constraints than amodal models. More importantly, since conceptual processing is based on re-enactments of feature maps, it more or less exhibits the temporal dynamics of original experiences. (Prinz & Barsalou 2000) Clearly, just as the temporality of stimuli-correlated representations ranges from the elementary scale (varying between 10 and 100 milliseconds) to the integration scale (varying between 0.5 to 3 seconds – corresponding to the experienced lived present) (Varela 1995; Varela et al. 2001), so offline simulations exhibit a temporality that spans from unconscious to more or less conscious processing. Importantly, even the multimodal phenomenal side of stimuli-correlated experiences inherently exhibits a temporal dimension.

4.5 A neurally embodied theory of language

Since it turned out that cognition is not quasi-linguistic, that language does not carry content if it is not processed within contextually embodied cognitive systems, and that the content of linguistic symbols is individuated in terms of context-sensitive phenomenal representations and simulations corresponding to a subject, which role plays language then? This very question is answered in what follows.

According to Pulvermüller's theory (2008), language is grounded in, and embodied by, action and perception mechanisms. The basic idea is that there are strong functional links between superior temporal speech perception, comprehension circuits (residing in perception- and action-related brain regions) and inferior frontal action control circuits. Pulvermüller et al. (2006) showed that distinct motor regions in the precentral gyrus which fired during articulatory movements of lips and tongue (syllables including *[p]* and *[t]* sounds), also fired when subjects listened to the lip- or tongue related phonemes. This neuroimaging corroborates the thesis of specific links between the phonological mechanisms for speech perception and production (phonological somatotopy). The fact that high-frequency cortical

responses (gamma-band) occurred in the case of meaningful words (“crocodile”), but not in the case of meaningless word-like items (“crocodile”), suggests that high-frequency responses are generated by memory networks. Such an explanation predicts that a memory network should amplify the cortical activity. The mismatch negativity (MMN) is a standard indicator of cognitive processes being elicited by auditory stimuli. The enhanced activity of the MMN of familiar linguistic tokens in contrast to meaningless but pronounceable ones has been confirmed comprehensively (e.g., Näätänen et al. 1997; Endrass et al. 2004; Pettigrew et al. 2004).

Interestingly, it was shown that lesions in the superior temporal (speech perception) or inferior frontal cortex (speech production) causing aphasia normally impair both speech production and comprehension (Pulvermüller et al. 1991). Moreover, Horwitz et al. (2004) probed that the links between superior temporal and inferior frontal language areas depend on the amount of contentful information that is carried by words. These strong functional links between speech perception, comprehension, and speech production circuits depict the embodiment of word forms and semantics.

Where and how is word meaning represented in the human brain? Hearing the word “bird”, for example, frequently together with certain visual perceptions, strengthens the connections between visual neurons (e.g. form and colour detectors in the primary cortex) and perisylvian language areas which process the word form. In the case of action words, neural networks in the motor, premotor, and prefrontal cortex responsible for action control are linked to those word form processing language areas. A number of neuroimaging studies have confirmed category-specific activation for the processing of action and visually-related words and concepts (e.g., Chao et al. 2001; Kiefer 2001; Cappa et al. 1998).

According to the idea of category-specific semantic processes, different kinds of word meaning reside in different parts of the brain. However, could it not be that the differential activation is explained in terms other than semantic? Since nouns often have more highly imageable content than verbs, while verbs tend to occur more frequently, any difference in brain activation could be brought about by an imageability-frequency dissociation.⁶³ What we therefore need is evidence of category-specific semantic activation at precisely those locations the brain-based sensorimotor semantics predicts. Somatotopic activations in action word processing provide this: Hauk et al. (2004) reported fMRI data demonstrating that tongue, finger, and foot movements lead to specific somatotopic activation patterns which are

⁶³ Moreover, predictions on where category-specific activation is to occur have not always been precise. Martin et al. (1996) showed, for example, that semantic information about colour and motion occurred ~2 centimetres anterior to the areas that respond maximally to colour and motion.

similar to activation patterns, while subjects silently read action words related to face, arm, and leg (whereby psycholinguistic variables, such as word frequency, length, and imageability were pretty much identical). Moreover, Tettamanti et al. (2005) showed that hearing action words embedded into spoken sentences (e.g., “The boy kicked the ball”) triggers the activation of specific body part representations.

Why is somatotopic activation semantic and not epiphenomenal in nature? Pulvermüller gives three criteria that are supposed to separate comprehension from epiphenomenal processing – immediacy, automaticity, and functional relevance. In order that motor areas contribute to semantic processing, their activation should take place immediately, that is, within the first 200 milliseconds after stimuli can be uniquely identified as an incoming word, since it is known that early lexical and semantic processing occurs around 100-200 milliseconds after critical stimulus information comes in (Sereno et al. 1998). Indeed, early activation differences between motor representations of the face and leg while silently reading face and leg words have been ascertained (Hauk et al. 2004). Automatic processing purports that hearing or seeing a word almost inevitably leads to comprehending its content, even if the subject does not intentionally attend to the stimulus. Here, too, the semantic role of somatotopic activation was confirmed: Subjects, distracted by spoken language input (face-, arm-, and leg-related action words) through watching a silent video film exhibited somatotopic activation (Näätänen et al. 2001; Pulvermüller, Shtyrov et al. 2005). Motor areas are only then functionally relevant, if changing the functional states of these motor systems is sufficient for measurable effects on the semantic processes. Pulvermüller, Hauk et al. (2005) showed that strong magnetic pulses that elicited muscle contractions in the right hand led to faster processing of arm words compared with leg words, whereby the reverse result occurred when transcranial magnetic stimulation (TMS) was employed on the cortical leg area. This underpins the specific influence of somatotopic activation on the understanding of action-related words.

How is language, understood as comprising word forms (phonological and orthographical) and semantics, processed in the brain? The widely distributed networks representing words ('word webs') become active in a discrete manner, that is, they are either active or inactive, and the full activation of a word web competes with other word webs. Assuming discrete activations is not to say that those representations are not context-specific. Other brain states in general, and other cognitive network activations in particular, form contexts that prime the way a word web fires. If you like, this neurally underpins Wittgenstein's (1953) idea of word

meaning as a family of similar context-dependent semantic feature sets.⁶⁴ Moreover, discrete representations allow the disambiguation of a semantically ambiguous word – the respective semantic context primes one of the semantic subassemblies.

I am now going to add some further considerations concerning language with reference to Barsalou (2008): Linguistic symbols are also modal symbols (as visually and auditorily perceived forms). They are schematic memories of perceived written or spoken words that are integrated into simulators. As simulators for words are established in memory, they get connected to simulators for phenomenal entities to which they refer. For example, whereas the simulator for the word “racing bicycle” becomes associated with the overall simulator for *racing bicycle*, other simulators for words become linked to simulators' subregions, such as the one for “handlebar tape” to the respective part. From the point where simulators for words are associated with simulators for concepts, simulators for words can control simulations. While reading a linguistic symbol, the associated content-bearing simulator becomes activated, thereby simulating a potential phenomenal entity and in this way providing a semantic interpretation (insofar as language symbols *index* simulations). Inversely, while producing speech within communication, the content-bearing simulation activates words and syntactic forms which, if spoken, function as guidelines for the semantic simulations of the hearer (in respect thereof, linguistic symbols *control* simulations).

Although word associations function as heuristics for correct conceptual performance (e.g., synonym judgements), they do not amount to deep conceptual processing, but rather are superficial.⁶⁵ Simulations provide the true comprehension, often being indicated by linguistic forms. Additionally, extensive interaction occurs between the simulation and the linguistic system: The initial activation of linguistic forms activates simulations. Subsequently, words that refer to the simulated space-time regions become active. Finally, simulators that conceptually interpret these regions are activated. Since people constantly hear linguistic tokens corresponding to phenomenally-perceived situations or simulated situations, the statistical structures of the two systems (frequencies of properties and relations between them) mirror each other.

I will now outline some evidence for mixtures of the linguistic and semantic system in higher-level cognitive behaviour. Firstly, in property verification, participants produce more

⁶⁴ Pulvermüller makes the proposal that the distributed neural assemblies themselves, as a whole (that is, the nearby simultaneous activation of cortical processing word forms and conceptual content) implement the linguistic binding instead of one central locus. The bilateral nature of neural degeneration of semantic dementia (Patterson & Hodges 2001) speaks against clinging to one focal binding area.

⁶⁵ As already shown (see 4.1), performing symbolic operations (such as predication or conceptual combination) merely on linguistic forms, like manipulating symbols in an unfamiliar language, is not sufficient for real understanding.

object-situation responses and fewer linguistic responses, because the task is more conceptual and their responses take longer periods into account. Furthermore, it is confirmed that linguistic responses should significantly precede object-situation responses. Besides, superordinates were produced quite early, whereas subordinates and ordinates were as slow as object-situation responses. These findings indicate that superordinates were linguistically processed, whereby subordinates and ordinates were conceptually simulated. Various assumptions were also corroborated by fMRI: The word association task primarily activated left-hemisphere language areas, especially Broca's area. The situation simulation task activated bilateral posterior areas, typically responsible for mental imagery. In property generation, conceptual processing activated both localised regions; activations of the words association localiser occurred earlier than the activations found in the situation localizer.

Secondly, consider evidence from research on property verification: As predicted by Solomon and Barsalou (2004), if information of the linguistic system is exclusively sufficient for adequate task performance, then participants will adopt the linguistic strategy. This strategy is sufficient if the words for true properties are related to the words for the target object (e.g., "WATERMELON-seed"), and if the words for false properties are unrelated to the target object (e.g., "LION-wire"). When, however, the words for the false properties are related to the target object (e.g., "BANANA-monkey"), participants will use conceptual simulations. These predictions are neurally corroborated: Related false trials showed the activity of the left fusiform area – a region that contributes to visual imagery. In the case of unrelated false trials, this area was not active.

Thirdly, to revisit the discussion on abstract concepts, consider the following evidence: For abstract concepts too, mixtures of language and semantic simulation systems do the representational work. As confirmed by Wilson et al. (2007), the linguistic system was not more active for abstract concepts than for concrete ones.⁶⁶

Chaffin (1997) found that high-frequency words often produced semantic responses that described events, whereas low-frequency words often generated linguistic responses – for example, synonyms and sound or orthographical similarities.⁶⁷ One explanation is that high-frequency words strongly activate situated simulations in the simulation systems and low-frequency words mainly activate linguistic tokens in the linguistic system. Due to the fact that low-frequency words have not been linked enough to experiences to activate the

⁶⁶ Other neuroimaging experiments had only confirmed the thesis that abstract concepts are exclusively linguistically represented because the tasks posed encouraged superficial linguistic processing.

⁶⁷ This thought of linguistic responses can be linked with the theory of convergence zones: For example, cross-modal CZs associate spoken (auditory) and written (visual) forms of "table", for instance (Simmons & Barsalou 2003, p. 16).

simulation of familiar situations, they are associated with linguistic information. Conversely, since high-frequency words activate event simulations well known from experience to support situated action, they are associated with pragmatic information.⁶⁸

Consider the complex interactions between linguistic and simulation processing again: The activation of linguistic forms leads to activation of simulations. In turn, these activate words that manipulate and describe the simulations. Generally, Barsalou assumes that the symbolic structures and symbolic operations result from the interactions between the language and the simulation system. There is evidence that shows that syntactic structures and affordances available from simulations are related (Glenberg & Robertson 2000).⁶⁹

In the following section, we deal with empirical evidence for concept empiricism.

4.6 Further empirical evidence for concept empiricism

With a view to making the whole notion of embodiment more intelligible and in order to justify PSS in particular and concept empiricism in general, we outline now further empirical evidence that suggest the truth of concept empiricism.

Importantly, the following evidence makes up the central justification for the truth of PSS and the inadequacy of amodal symbols systems accounts. Purely behavioural and neurobehavioural evidence is outlined.

(1) *Occlusion effects* during property generation: Wu and Barsalou (2005) induced half of the participants to generate properties for the noun concept [LAWN], and the other half for the same noun by a modifier (ROLLED-UP LAWN). Decisively, the participants of the first half hardly generated internal properties of [LAWN], such as DIRT and ROOTS, whereas the second half generated them much more frequently. What does this speak for? According to amodal theories, the meaning of the compounded linguistic token ROLLED-UP LAWN is a function of the meaning of its components. Here more importantly, without ad hoc

⁶⁸ Moreover, Following the principles of content addressability and encoding specificity, depending on the kind of cue (sensory-motor (a picture, for instance) or linguistic), the conceptual system activates faster either simulations or linguistic tokens respectively. This may be considered as another variable that has to be taken into account in order to explain the variations in activations and time.

⁶⁹ The basic ideas involved are reflected in Glenberg and Robertson's indexical hypothesis (1999, 2002). This thesis says that language comprehension consists of three processes: First, phrases are indexed to actual phenomenal objects or corresponding perceptual symbols. Second, by means of the indexed phenomenal object or perceptual symbols (which are integrated into a frame) affordances are derived (e.g., sentences that contain "chair" are understood by deriving 'you can sit on'). Third, the affordances are meshed guided by the sentence's syntax (e.g., to understand the sentence "John sits on the chair" one has to mesh the affordances of the chair and of John, so that John is on and not under the chair).

assumptions the accessibility of DIRT and ROOTS as properties for [LAWN] should not vary by adding the modifier ROLLED-UP, according to amodal theories. Yet, exactly the opposite occurred. Modal theories explain this in terms of internal simulations – the first half simulated the lawn's surface, whereas the second half simulated a rolled-up lawn, thereby accessing more internal properties than the first half. It was ensured that the modifier ROLLED-UP did not increase the number of occluded properties at every word. Furthermore, this finding also appeared both for familiar (e.g., HALF WATERMELON) and for novel noun combinations (e.g., GLASS CAR).

(2) *Size effects* during property verification: During property verification, the data in need of explanation are error rates and response times. Critically, the larger the property, the longer the verification times. This indicates that the respective properties to be verified are perceptually simulated. Moreover, scans of participants' performance via fMRI show activation in the left fusiform gyrus – a region often active in mental imagery and high-level object perception.

(3) *Shape effects* during property verification: Solomon and Barsalou (2001) examined whether verifying a property (MANE) for one concept (LION) facilitates verifying the same property for another concept (PONY). If MANE is represented by a single amodal symbol, abstracting over variously shaped manes, processing MANE for [LION] should facilitate subsequently processing MANE for [HORSE]. If, however, participants simulate manes during verification, simulating a lion mane should not benefit simulating a pony mane, because lion and pony manes have different shapes. Accordingly, if participants successively verify MANE for [PONY] and MANE for [HORSE], the former should facilitate the latter simulation. Not surprisingly, simulating MANE for [PONY] was *only* facilitated by previously simulating MANE for [HORSE]. By the way, it was ensured that the effect did not arise from the greater similarity between [HORSE] and [PONY] than between [LION] and [PONY], the property was the critical factor.

(4) *Modality switching* during property verification: Pecher et al. (2003) showed that, for instance, verifying LOUD for [BLENDER] worked faster when RUSTLING was previously verified for [LEAVES] than when TART was verified for [CRANBERRIES]. Corresponding to modality switching during perception, the comparative delay is explained in terms of modality shifts during simulation. Probably, selective attention has to shift from one to another modality. Obviously, it is not the sole explanation – perhaps properties from the same modality are more closely associated among themselves than properties from different modalities, thereby priming across verifications trials. Yet, when checked, neither properties

from the same modality were more associated than one from different modalities, nor were highly associated properties verified faster than unassociated properties.

(5) *Shape and orientation* during language comprehension: Zwaan, Stanfield, and Yaxley (2002) stated the influence of coincidence/anti-coincidence between reading a vignette and the response time subjects take to name objects of shown pictures, especially in respect of the object's shape. Using the example, if a bird with outstretched (or folded) wings was shown and the participants previously read sentences that described a flying (or sitting) bird, they designated objects faster than when the sentences' implicit shape did match the pictures' shape.

Furthermore, Stanfield and Zwaan (2002) demonstrated that the orientation of objects affects language comprehension. For instance, some participants read sentences about someone pounding a nail into the wall, whereas others read sentences about someone pounding a nail into the floor. Directly afterwards, participants saw a picture of an isolated object and had to show whether it had been named in the sentences. If the orientation between sentences and pictures (in both a horizontal nail or in both a vertical nail) coincided, the verification was faster than the one in which orientation between sentences and picture differed. In conclusion, both experiments ought to show that people simulate objects during sentence comprehension.

(6) *Movement direction* in language comprehension: Glenberg and Kaschak (2002) intended to demonstrate that people's understanding of sentences that describe actions are based on simulating the actions in their motor systems. They ascertained a correlation between the coincidence of the direction of the press button movement (whereby participants indicated whether sentences are grammatically correct/incorrect) and the direction of the described action and the response times. Using the example, the sentences "Open the drawer" and "Close the drawer" were verified faster, if the former was verified via pressing the button towards their bodies and the latter with a button pressed away from their bodies than when the described action and the indicating movement were contrary.

(7) *Category-specific deficits*: Since visual processing is significant for interacting with [LIVING THINGS], such as [MAMMALS], and, according to modal theories, knowledge of those categories resides not insignificantly in areas responsible for visual processing, lesions in visual areas should increase the probability of losing knowledge of those categories. Or, because action is central for interacting with [MANIPULABLE OBJECTS], such as [TOOLS], and, according to modal theories, knowledge of those categories resides to a great extent in motor areas, damage to motor areas should increase the probability of losing

knowledge of those categories. Both cases are depicted in Damasio and Damasio (1994) or in Humphreys and Forde (2001). Analogously, lesions in areas responsible for colour processing cause colour knowledge deficits, damage to areas responsible for space lead to deficits in knowledge of location.

(8) *Neuroimaging studies of category knowledge*: Processing categories that strongly depend on visual information, such as [ANIMALS], shows considerable activation in visual areas. Analogously, handling categories that heavily depend on action, such as [TOOLS], displays striking activation in the motor systems (Martin et al. 1996). The analogue was also found for processing colour categories (Shao & Martin 1999). More concretely, while participants looked at manipulable objects in isolation (for example, a hammer) lying in an fMRI scanner, a brain circuit that underlies the grasping of manipulable objects became active (besides, a clear confirmation of sensorimotor coupling). It is worth noting that while participants were viewing animals, buildings, and faces, this brain circuit did not become active. Since the participants neither moved nor viewed pictures of others' actions (to exclude the causal efficacy of mirror neurons), Shao and Martin concluded that the circuit's activation constitutes a motor inference which includes the way to act on the perceived object. Similarly, Simmons, Martin, and Barsalou (2005) induced participants to view food pictures in an fMRI scanner, whereby such brain areas became active as represent the taste of food, elsewhere being activated during ingestion. They construe this to indicate that as participants viewed food, category knowledge became active which generated taste inferences by use of simulations in the gustatory systems.

To what extent do these experiments determine the empirical or philosophical theories? In any case to the effect that knowledge resides in sensorimotor areas. Because those areas generate phenomenal experiences in online processing, and we have no phenomenal experience of amodal symbols, I would even go so far as to say that those experiments also entail that there are no amodal symbols.⁷⁰

Now, let us consider evidence that shows how subjects make use of situated conceptualizations consisting of four types of situated inferences: Inferences about (i) the goal-relevant properties of the pertinent category, (ii) the background properties, (iii) the appropriate actions for achieving the focal goal and (iv) the phenomenally opaque states that the agent will probably have to go through while interacting with the category (e.g. evaluations, emotions, cognitive operations).

⁷⁰ This is believed to anticipate the objection that empirical evidence underdetermines the truth of the respective theories.

(1) *Inferences about goal-relevant properties of the pertinent category*: Barsalou (1982) demonstrated that the time participants take to verify a property for a noun depends on the context that is indexed by the sentences they previously read which contained the focal noun. Using the example, participants verified FLOATS 145 ms faster after reading the sentence, “The basketball was used when the boat sank”, than after reading “The basketball was well worn from much use.” Thus, the concept for [BASKETBALL] did not produce the same representations across both contexts. Similarly, context effects on word encoding within the frame of memory were shown, moreover context effects on lexical access during sentence processing. Hence, a category is not represented by one general description (not to mention individually necessary and jointly sufficient conditions) that functions across different situations, but rather tailored to contextual requirements.

(2) *Inferences about background settings*: The decisive point is that instead of being represented isolated, categories should be accessed situationally embedded. Vallée-Tourangeau et al. (1998) induced participants to give instances of common taxonomic categories (using the example, [FRUIT]) and of ad hoc categories ([THINGS PEOPLE TAKE TO A WEDDING]). Afterwards, they inquired about the generation strategy that participants had used, the choice being that instances came to mind automatically/unmediated, via semantic taxonomy or by means of retrieval from experienced situations. More than half of participants reported the experiential strategy. Quite possibly, this experiment can be understood as checking the self-conception that the respondents had.⁷¹ The experiments described earlier also showed the situational embeddedness of category representation (e.g., to generate properties for [WATERMELON], participants inadvertently produced setting properties, such as PARC or PICNIC TABLE), thereby not relying on introspection.

(3) *Inferences about actions the agent could take to reach an associated goal*: This is to show that the situated conceptualizations of categories place the conceptualizer in the represented situations. Reading a sentence about an action (without mentioning an agent) activates a motor representation of it (Glenberg & Kaschak 2002). As already mentioned, the grasping circuit was activated when participants looked at manipulable artifacts in isolation (Chao & Martin 2000); similarly, merely reading an action word activated the respective part of the motor strip (Hauk et al. 2004). As was intended, these findings demonstrate the action-

⁷¹ This may be considered as an objection to the view that the classical folk psychology reflects the self-conception of the participants. Clearly, if claims about self-conceptions are not the result of empirical studies, they cannot be count as justified. Apart from that, one might challenge the relevance of self-conceptions for a theoretical understanding of mental processes.

orientation of the conceptual system by priming relevant actions in the motor system; the activation of a concept prepares the conceptualizer for interacting with its instances. There is further evidence in social cognition: While viewing other's faces to categorize their emotional states, participants simulated the respective emotional expressions on their own faces (Wallbott 1991). Analogously, banning the simulation of other's emotional states decreased their ability to categorize (Niedenthal et al. 2001); or the accuracy of participants' categorizations correlated with the extent to which their facial simulations were recognizable (they were videotaped). To this extent, simulating others' emotional expressions is regarded as motor inference of the conceptual system in order to support situated action – usually, it is useful for the perceiver to adopt the same emotional state as the perceived person.

(4) Inferences about mental states, that are phenomenally regarded as internal, which the conceptualizer will have while interacting with instances of the conceptualized category: As participants simulated particular situations with the category [WATERMELON], in addition to background properties, they also generated properties about presumptive internal states that they should experience in imagined situations, such as evaluations whether the objects are good/bad, effective/ineffective, emotional reactions to objects, such as happiness or fear, and cognitive operations relevant for interacting with the respective objects, such as comparing the object to alternatives (Wu & Barsalou 2005).

While we dealt with empirical evidence for concept empiricism in this section, the next one revisits mechanistic explanations.

4.7 Identifying PSS as a mechanistic explanation

The aim of the following section is to make a proposal on how PSS could be conceived of as a mechanistic explanation. As we have argued for the rightness of mechanistic explanations (see 3.2), and since we adhere to PSS, this intention is natural.

The empirical evidence is provided by both bottom-up (lesions correlated by conceptual deficits) and top-down (fMRI studies of conceptual tasks) experiments. It thus also seems appropriate to say that PSS takes several levels into consideration – at least two functionally individuated brain regions, and correlated behaviour of the cognitive system as a whole. I think that one could divide PSS into the following three levels: very high-level functional account, high-level functional account, and the activity of brain regions. Very high-level functional properties are productivity or variable embodiment, for example. High-level functional properties are the frame structure, the possession of simulators, or filling-ins, for

example. Simulators, for example, are controlled by cross-modal and modality CZs residing in the perirhinal cortex; entorhinal cortex, amygdala, and hippocampus, respectively (cf. Simmons & Barsalou 2003, pp. 471-2).

Clearly, PSS is a how explanation – the re-enactment of sensorimotor areas explains why we are able to think about things in their environmental decoupling, or why our thoughts are often accompanied by phenomenal imagery.

Moreover, the re-enactment of feature maps of sensorimotor areas constitutes understanding or offline cognition (interlevel constitution). An example of an intralevel (at high-functional level) causal relation is the ongoing alternation of stored frames brought forth by selectively attending to phenomenal representations of the environment.

Consider an example that illustrates in what way PSS might be productively continuous: Imagine that a friend tells you that the ceiling in his cellar starts going mouldy. Thereupon you screw up your nose and you reply that it has to smell his cellar. How can your capacity to add an emotion (disgust) and an olfactory feature to the talking (what is supposed to stand for the very high-level functional property of productivity) be explained? Your brain could make use of CZs occurring in vision to categorize CELLAR, CEILING, and MOULDY. More or less parallel to this cross-modal CZs, that capture the correlation between visual and olfactory properties, could re-activate the feature maps constituting the disgust. Approximately in parallel with this feeling CZs that reside in somatosensory areas in the right hemisphere link your disgust to your facial expression (cf. Adolphs et al. 2000). Lastly, this complex bodily simulation (what is supposed to stand for the high-level functional property of a simulation) feed activation to the linguistic response systems.

4.8 Problems for the postulation of amodal symbols

Now we are going to look at objections that exclusively take issue with amodality.

Perceptual states of the perception system are transduced into a new representational system whose symbols bear no structural similarity with the entities of the perceptual states that produced them. No structural similarity means that the amodal symbols do not retain the systematic relation between the phenomenal perceptual states to which they refer (e.g., the amodal symbols “red” and “orange” are not more similar to each other than “red” and “blue”).

What are the problems for amodal theories? First, they are unfalsifiable. All conceivable empirical evidence is explainable post hoc in terms of amodal symbols. By means of ad hoc

assumptions they can explain every possible behavioral and neural evidence.⁷² Second, the thesis that concepts are amodally grounded lacks direct evidence. There is no direct evidence that amodal symbols are causally relevant for conceptual processing. Third, although theories of amodal symbols are indirectly confirmed by their capacity to implement conceptual systems (productivity, type-token distinction, systematicity, propositional structure), they cannot implement all computational functions – such as spatio-temporal knowledge (Clark 1997). Moreover, as we have seen, modal symbols can implement computational functions too. Since in addition to that modal theories are empirically well confirmed, the justificatory force of implementing conceptual functions by amodal symbols is quite debatable. Fourth, if it is assumed that phenomenally perceivable things count among the contents of offline thoughts and that these contents are created in online cognition by our sensorimotor systems and if offline thought is about phenomenally perceivable things and makes use of arbitrary representations, then the representational format being constructed by our sensorimotor systems has to be transduced into the format of arbitrary representations. This transduction is problematic, however, for the following reason: Because pervasive interactions between higher- and lower-level types of mental processes exist, a substantial additional cost were required in comparison to the possibility that higher-level cognition use the same representational format as lower-level types cognition. Fifth, as already shown, they also do not satisfactorily answer the symbol grounding problem – how amodal symbols get their contents within individual cognitive systems. Sixth, a related problem is how human beings are able to understand sentences in the absence of their phenomenal referents, if the processing of amodal symbols is exclusively syntactical. There is no doubt that people do comprehend such sentences. But if nothing in their environment can be responsible for the symbols' grounding, how do they understand them? Seventh, one solution to the last three problems reveals another weakness in the theories of amodal symbol systems. Namely, one could assume mediating perceptual representations (Harnard 1987). Accordingly, perceptual representations are assigned to amodal symbols in long-term memory (e.g., CAT is associated with perceptual memories of cats). Since the perceptions of cats activates perceptual memories of cats, which activate CAT as amodal symbol, the transduction problem is solved. During symbol grounding, the inverted course of activation is the case. Now to the problem. If sensorimotor memories and speech perception and production areas

⁷² It is no wonder that they are unfalsifiable because such theories take neither phenomenological nor neural nor neurophenomenological constraints seriously. As one can model or explain in principle all conceivable behavioural evidence in terms of representations of amodal symbols and map them onto all conceivable neural activations, such theories cannot be wrong.

do all the work, why additionally postulate amodal symbols? Are they anything more than redundant?

The subsequent section anew provides indirect reason for concept empiricism by challenging the objections being made to it.

4.9 Discussing objections to concept empiricism

In the following I am going to discuss several objections to PSS in particular, or to perceptual symbol systems altogether. The objections stem from, or are noted in (i) Weiskopf (2007, forthcoming), (ii) Machery (2007), (iii) open peer commentaries on Barsalou's "Perceptual Symbol Systems" (1999), and (iv) Siewert (1998) and Strawson (1994).

(i) (1) If the vehicles of thought are the vehicles of perception, how can thinking be separated from perceiving? (cf., p. 6) Let me call on several points that make the separation clear: Perceptual presentations and representations comprise non-attended contents (in other words, perceptual symbols are schematic in consequence of selective attention); conceptual processing relies more on memory than perception; a common set of sensorimotor contents is used by different systems; CZs initiate the re-enactment of feature maps; whereas feature maps code modality-specific information in order of perception, analytic and holistic CZs conjoin features to conceptual properties; besides, in capturing the correlations between various analytic and holistic properties, modality CZs are constitutive of category representations; as cross-modal CZs concentrate on statistical correlations between the six modalities, they are far from being mere bottom-up perceptions; since PSS exhibits productivity, it can simulate non-experienced entities.

(2) According to Weiskopf, convergence zones satisfy the necessary and jointly sufficient conditions for being representations: They are endogenously deployable and causally implemented in categorization and inference. To this extent they are amodal representations and from this he concludes that not all concepts are completely composed of percepts (strong global empiricism). Weiskopf refers to a strategy the concept empiricist could pursue to elude this objection: Convergence zones are not representations at all, but mechanisms for controlling representations. Let me shortly elaborate this answer of the concept empiricist: Remember (see 4.2) Simmons and Barsalou's suggestion that the conjunctive neurons of convergence zones only function as stand-alone representations, if they feed forward automatically to linguistic responses under very routinised conditions, such as categorization or word associations. Nevertheless, non-automatised processing requires the activation of

feature maps. (2003, p. 456) Subsequently, Weiskopf alleges three different reasons for rejecting this reaction of the empiricist. I think that the distinction between representational system and controlling mechanism is precisely right under conditions that are not highly routinised, and therefore I am now going to contest the three reasons being adduced by Weiskopf. Firstly, he claims that systems could contain mechanisms in virtue of implicitly representing rules, such as the modus ponens. As we have already seen, referring to a logical syntax as the scaffold of thought is empirically unjustified. Secondly, he says that neural mechanisms themselves can contain explicit representations (e.g., the amodal representation CAT tokened in a convergence zone re-enacts lower sensory regions). As evidence for this claim he refers to the mechanistic framework – that is, CAT is an example for the notion of a 'part'. To take only a single point again – it seems quite flawed and ad hoc to regard an arbitrary linguistic symbol that occurs at best in a mechanism's whole behaviour as a causally efficient part which participates in the generation of the whole behaviour. It is a match for the 'fallacy of phenomenological reification'. Thirdly, he claims that nomic co-variation is sufficient therefore that convergence zones are representational (that is, the function reliably brings about perceptual representations). He further says that if one adds the condition of playing the role of 'standing in' for absent entities or sensorimotor experiences in the first place to the condition of nomic co-variation (because it is too wide as an individual condition) in order to be sufficient for representation, one excludes not only convergence zones from being representational but also motor representations, and this seems to be untenable. (2007, pp. 11-4) Why should motor representations be excluded? Somatotopic activation patterns are nice examples that they are not.

(3) Consider the following example of an action compatibility effect: Participants responded faster when the direction they took to press the button corresponded with the direction of the movement described. For example, when reading the sentence “John gave me the back bag”, they were faster to move the response key towards themselves than away. Glenberg and Kaschak (2004) explain this in terms of the influence between action-oriented motor representations that ground sentence understanding and the corresponding execution of bodily movements. If the directions of representation and movement correspond, the execution of the bodily movement is facilitated, if not, the action-oriented representation interferes with the execution. Weiskopf correctly criticizes that the compatibility effect in the example “John gave me the back bag” cannot be explained by corresponding directions, since the reader's action-oriented representation of its own action consists in reaching out his arms to take the back bag. Borregine and Kaschak (2006) provide another explanation:

Reading a sentence brings about a construction of an enactive simulation of perceiving a situation whereby selective attention is paid to the giver. These representations comprise entities that are transferred from one place to another. Since perceiving and acting have a common representational code, corresponding directions of simulated transfer (simulating the motor behaviour of the perceived giver) and executed motor movements (pressing the button) facilitate the latter.

(4) Weiskopf argues for the thesis that linguistic understanding does not require enactive simulation. Knowing the truth condition of a sentence is sufficient for comprehending it. This knowledge implies being able to draw certain inferences (e.g., “John wears glasses” entails “Someone wears glasses”). He claims further that the decisive point is whether these inferences comprise knowledge of appearances and affordances of the described objects. He finally says that syntactic and semantic features of sentences are available to anyone without being committed to being able to enactive simulation. (forthcoming, pp. 15-7) It seems to me that his objection is unacceptable. How is the knowledge of truth conditions implemented? Is it not in the form of offline simulations? What are the semantic and syntactic features independent of simulations? If one really draws such inferences without simulating, is it conceptual processing at all? Is it more than playing meaningless language games? And does such a game not consist of simulating written or spoken words? How are inferences actually represented? There is a lot of evidence that inferences reside in sensorimotor simulations (cf. Schmalhofer 2007). Furthermore, mastering inferential linguistic behaviour free of conceptual processing is no objection to concept empiricism, because it can be simulated by making use of the superior temporal (speech perception) or inferior frontal cortex (speech production).

(5) Simulations are not fine-grained enough to distinguish between certain sentences that have different truth conditions: For example, the enactive simulations of “A man stood in the corner” and “A man waited in the corner” cannot account for the different truth conditions of both sentences (Weiskopf, forthcoming, pp. 17-8). I think this objection is indicative of a widespread misunderstanding of perceptual symbol systems: What could be the difference between the two sentences if one understands simulation as an isolated static picture of a man in a corner that appears in the mind's eye? I think that because in perceptual symbol systems emotional states also belong to concepts, for example, waiting and standing could easily be differentiated by simulating the different emotional states that accompany waiting and standing.

(6) Making inferences about the truth conditions of sentences that deal with novel lexical items indicates that visualization is not constitutive of understanding (e.g., inferring from “the dax was sleeping on a bed” that the dax is a living creature, since only living things can sleep) (ibid., p. 18). It is neither a matter of conscious visualization (that is, re-enactment of multimodal sensorimotor areas does not have to be conscious), nor does such a case speak against simulation. How do we understand SLEEPING and LIVING if not in the form of more or less conscious simulations?

(7) The argument from content-vehicle conflation: Affordance compatibility effects, for example, only show that our psychological states carry content about affordances in linguistic understanding. It by no means shows that this content is carried out by the same vehicles that are responsible for sensorimotor processing. In other words, the fundamental point is that processing content about affordances or visual properties does not imply that the processing vehicles are themselves sensorimotor. Additionally, the affordance compatibility effects do not even show that information about affordances is perceptual and not conceptualized content (reasoning, thinking). (ibid., pp. 20-1) Already mentioned neuroimaging studies show the utilization of sensorimotor vehicles. The latter point is quite weak but not unusual. The point of perceptual symbol systems is precisely that a sensorimotor system can implement a fully functional conceptual system.

(8) The frame problem arises for amodal *and* embodied approaches. The reason for this is that the frame problem is quite independent of the choice of representational vehicles. (ibid., p. 29) Actually, the contrary is the case. Bearing in mind that contents and vehicles are inherently connected and that higher-level cognition (including action planning) makes use of sensorimotor vehicles and contents, a functioning link between higher-level cognition and action is readily possible. To solve the frame problem, the decisive point is to recognize the wide range of subpersonal processes working in parallel (motor processing of the dorsal stream; non-phenomenal simulations of the forward model; situated conceptualizations) that manage the contextual selection of action-relevant contents. Moreover, since perception and conception make use of the same sensorimotor structures, resemblance relations make the selection of action-relevant content particularly easy. Moreover, the basic idea of the global workspace hypothesis (Baars 1988, 1997) is that multiple parallel processes are responsible for the selection of relevant information. Several parallel specialist processes (such as those responsible for language production or some aspects of perception) compete and cooperate for access to the global workspace. If some information has accessed the global workspace, it is sent back to all specialist processes. Precisely because of the interplay between parallel

selection and despatch back to the multiple parallel processes, consciousness is cognitively or causally effective.

(ii) (1) It is one thing to claim that re-enacted feature maps are tailored to the respective current context; it is quite another thing to determine the mechanisms that successfully enable context-specificity of re-enacted feature maps. Context-sensitivity does not come for free for perceptual symbol systems. (2007, p. 12) As we said above, resemblance relations, the forward model, non-phenomenal vision and situated conceptualizations are good candidates for those processes.

(2) Some proponents of the amodal approach (e.g., Fodor 1975) acknowledge that perceptual simulation plays a role in conceptual tasks. Thus perceptual simulation is not suitable for distinguishing neo-empiricist and amodal models. (2007, pp. 17-24) I think that drawing on amodal quasi-linguistic propositions that are supposed to be causally efficient is incompatible with embodied cognition.

(3) There is a lot of behavioral and neuropsychological evidence for amodal and analogical representations of cardinality. (2007, p. 37) This objection cannot be rejected completely. What can be held against it is that Lakoff & Núñez (2000) made a proposal on how the embodied mind processes mathematics.

(iii) (1) The fact that lesions in particular sensorimotor areas are sufficient for knowledge deficits which depend on these areas (e.g., knowledge of BIRDS depends highly on visual processing) is consistent with claiming that amodal symbols represent concepts, and that amodal symbols reside in sensorimotor areas (Adams & Campbell 1999; Aydede 1999). Behavioral findings, however, tell another story: Occlusion effects during property generation or modality switching effects are neither predicted nor adequately explained in terms of amodal symbols. (Barsalou 1999, p. 637) Moreover, the approach of amodal symbols suffers from the frame, the symbol grounding (ibid., p. 638) and the symbol transduction problem. Postulating amodal symbols as supposedly residing in sensorimotor areas entirely neglects interlevel constraints on multilevel explanations (cf. Craver 2009, p. 249).

(2) Could it not be that CZs stand in for feature maps during symbolic activity, thereby processing something like amodal symbols? Could it not be that sensorimotor processing is ultimately epiphenomena? (Adams & Campbell 1999, p. 610) Aside from the fact that Damasio (1989) assertively says that the CZs do not stand in for full-blooded conceptual sensorimotor simulations in the absence of re-enacted feature maps, if this (amodal symbols in association areas) were the case, how did the above-mentioned behavioral evidence come

into existence? What is needed, of course, to clarify the matter definitely, are neuroimaging studies that can determine the activity of feature maps and association areas during perceptual and conceptual processing. If it were then the case that lesions in specific sensorimotor areas separated from local association areas would be sufficient for corresponding knowledge deficits, sensorimotor simulations could not be mere epiphenomena. Clearly, if one claimed that concepts neither reside in sensorimotor areas nor in adjacent association areas and that insofar perceptual simulations were epiphenomena, then lesions in particular sensorimotor areas could not be sufficient for knowledge deficits that depend on those sensorimotor areas.

(3) Landauer (1999) takes Barsalou to mean that human-like knowledge cannot be learned or represented without human bodies (that is, a silicon-based system cannot acquire human knowledge). Landauer shows, however, that computers implementing latent semantic analysis (LSA) emulate humans on several knowledge-based tasks. Yet Barsalou did not exclude that amodal symbols could emulate humans in behavioral forms. All that he claimed was that if knowledge is grounded in sensorimotor areas – and compelling evidence of this exists – then humans and computers represent knowledge differently, since their sensorimotor systems differ. Conversely, if future computers could have sensorimotor systems closer to ours, quite possibly they could represent knowledge similar to humans. (Barsalou 1999, p. 639) Moreover, empirical evidence shows that knowledge does not consist of word co-occurrences: Aphasics lose language without losing knowledge (Lowenthal 1992); affordances influence language comprehension that cannot be explained in terms of word co-occurrence (e.g. ad hoc categories) (Glenberg et al. 1987). To anticipate the objection that affordances or perceptual variables are merely correlational, not causal, it was shown that non-linguistic factors significantly influenced behaviour in a task that involved only linguistic stimuli, when linguistic factors remained constant (cf. Solomon & Barsalou 1999a, 1999b).

(4) Oehlmann (1999) confronts PSS with the problem of explaining how people can know that they know a solution to a problem without the need to simulate the entire solution. Barsalou gives the following explanation: While simulating a solution, only the initial and final states are selectively attended to and therefore stored in working memory. One then switches back and forth between initial and final states and a shortened version of the simulation becomes associated with the complete simulation. If one later perceives or simulates the initial conditions, the shortened simulation is sufficient for producing a

internally or externally spoken response. Perhaps this objection is motivated by wrongly regarding PSS as a recording system. (Barsalou 1999, p. 645)

(5) Abstract concepts: (i) Landauer (1999) and Ohlsson (1999) regard perception as entirely improper for representing abstract concepts. To decide, however, whether PSS or any other approach can account for abstract concepts, it is necessary to identify the contents of those cognitive representations. (Barsalou 1999, p. 664) Without a cogent individuation of their contents, how can one claim to understand them and that this understanding cannot go back to perceptual structures? (ii) How is it possible that people understand abstract concepts of non-existent entities (such as *the end of time* or *electromagnetic field*), if those entities cannot be experienced? (Ohlsson 1999; Toomela 1999) Very simply, because PSS exhibits productivity – extracted components of past experiences are combined to produce novel simulations. *The end of time*, for example, could be understood via simulating the ends of familiar processes, like *the end of a journey*, and applying it to a simulation of *time* that could consist of a clock which stops running. Moreover, it is worth noting that abstract concepts do not come for free for amodal symbol systems. Why should non-experienced entities be easier for amodal theories to handle than for perceptual symbol systems? (Barsalou 1999, p. 647) And, as already mentioned, to decide whether they can handle them easier or not, one needs an individuation of the content of those concepts. And where is one supposed to search for them if not in transparent representations of the environment or internal processes.

(iii) Ohlsson (1999) criticizes Barsalou's notion of falsity: The lack of fit is not sufficient for the falsity of a simulation. Or, to put it more trenchantly, the absence of evidence is not evidence for the absence. For example, imagine seeing a book on a table. While you are turned away, someone removes it. According to Barsalou's account – as the objection goes – if you turn around again, your simulation is false. What is wrong with this objection? Crucially, a simulation is *purported to be* about a perceived situation. Of course, some simulations are only false if one failed to find a fit in any relevant situation – naturally, advocates of PSS do not want to conclude from failing to see Martians in one situation that they do not exist. (Barsalou 1999, p. 648)

(6) According to Ohlsson (1999, p. 630-1), Barsalou wrongly confuses selection with abstraction – selectively storing handlebars while perceiving a bicycle is tantamount to creating an abstraction of bicycles at large. Yet another easily refutable objection. Barsalou approximately identifies selection with schematization (abstracting a focal content from a perceived situation). He is by no means assuming that schematization is abstraction, or that schematization is sufficient for having concepts. Something only counts as a concept when

many schematic memories have been integrated into a simulator. Whereby he never focused on abstraction per se. If he were depicting abstraction with regard to PSS, he would define it as the deployment of a simulator that re-enacts an abundance of experiential forms. (Barsalou 1999, p. 642)

(iv) Finally, let me present an objection that is written large in a serious misunderstanding of embodied simulation. The fact that you experience hearing a sentence you understand differently from hearing a foreign one you do not understand, though the imagery (auralizing each sentence) is nearly identical in both cases, suggests that the experiential difference cannot be explained in terms of imagery. What if one realizes that the multimodal re-enactments of feature maps – that is, imagery – differs decisively from one another? It seems obvious that this objection is motivated by wrongly considering imagery as nothing more than hearing, speaking or seeing meaningless linguistic tokens, like the misapprehension of imagery in terms of static pictures that have to be read by a central processor to generate content at all. As we have seen, according to Pulvermüller's theory of language, the phenomenal difference amounts to drawing on sensorimotor contents stored in memory in the case of understanding, as distinct from hearing a foreign sentence. If you like, processing word forms and processing content bearing sensorimotor simulations as a package constitutes mental imagery.

Because we have presented a lot of empirical evidence and formulated a theory of the mind against the backdrop of this evidence, we are now able to consider which of the available models of the mental is most adequate to grasp the causally efficient processes being determined.

5 Against computationalism (both symbolic and connectionist)

The aim of the following chapter is to present core properties of three different models of mental processing. I will argue that two of them are inappropriate in several respects. It is important to point out that what I want to provide is at best an intuitive, non-mathematical understanding of the non-computational dynamical model that is hereby endorsed.

5.1 A critique of symbolism

It is important to make clear that the Computational Theory of Mind (CTM) grasps the mind as a digital computer in a literal sense. It is a much stronger claim than merely seeing

computation as a modelling technique for mental processes. Fundamentally, not each process that is computable is directly equal to a computation.⁷³

The Representational Theory of Mind (RTM) says the following: Intentional states such as beliefs and desires are relations between certain propositional attitudes and language-like symbolic representations (the contents of propositional attitudes). For example, hoping and believing that the table is green are different propositional attitudes with the same semantic value, namely, that *the table is green*. Different propositional attitudes take on different functional relations to symbolic representations. (Fodor 1975)

According to the Computational Theory of Reasoning (CTR), symbolic representations have both semantic and syntactical properties, whereby reasoning processes are only sensitive to the syntax of the symbolic representations ('structure sensitivity'). Symbolic representations are manipulated according to rules which operate only on the representations' shapes, which are arbitrarily related to their potential contents. Due to the fact that mental representations have a syntactical and semantic combinatorial structure, so Pylyshyn, classical operations can apply to them by referring to their form. Accordingly, transforming one mental representation into another one happens by logical rules, such as transforming the representation of the form P and Q into P, or the representations of the forms P or Q and not-P into Q.

The interpretation of symbolic representations is not intrinsic to the computational system. Therefore computation counts as formal symbol manipulation. That is not to say that computational manipulations do not have to be interpretable, on the contrary. More precisely, all interpretations of symbols and manipulations have to be aligned with another (Fodor & Pylyshyn 1988).

The CTM aims at showing that 'intentional realism' and the assumption that all mental processes are causal processes in need of mechanistic specification are compatible. Intentional realism purports that mental states have semantic properties and that mental states are causally relevant for behaviour. 'Formalization' and 'computation' are the decisive technical concepts. Formalization denotes how semantic features of symbolic representations can be encoded in syntax-based inference rules that are independent of semantics. In doing so, the semantic values can be processed in a manner that is only sensitive to syntax, without relying on a reasoner who must use semantic intuitions. Thus, formalization depicts how semantics can be linked to syntax. Since any processes that are only sensitive to syntax can be mechanically duplicated (Turing's computing machine), such mechanisms can assess any

⁷³ The CTM also has to be distinguished from viewing the software-hardware distinction characteristic of computers merely as a guiding metaphor for grasping some properties of the mind.

formalizable function. Thus, computation shows how syntax can be linked to causation. Since causation and semantics are both connected to syntax, causation and semantics are linked with each other. Reasoning processes that factor in the representations' semantic values can be executed purely mechanically, namely, by having a 'syntactic engine' that tracks all semantic properties by corresponding syntactical properties and is causally efficacious as regards reasoning.

Fodor (1980) and Pylyshyn (1980, 1986) (1988) further claim that the symbolic level, as a natural functional level of its own, is independent of physical implementation.

Succinctly, computation is implementation-independent, systematically interpretable formal symbol manipulation.

Furthermore, Fodor and Pylyshyn (1988, p. 13) claim that both the semantic and syntactical combinatorial structures have physical counterparts in the brain:

[...] the symbol structures in a Classical model are assumed to correspond to real physical structures in the brain and the combinatorial structure of a representation is supposed to have a counterpart in structural relations among physical properties of the brain. For example, the relation "part of," which holds between a relatively simple symbol and a more complex one, is assumed to correspond to some physical relation among brain states [...] (1988, p. 13)

Mapping quasi-linguistic structures (structurally atomic concepts or structurally molecular propositions) onto the brain with a view to determine causally efficient structures was called *conceptual* and *propositional modularity*, respectively, by Stich (1983) and Clark (1993, pp. 190-214). First, let me give an example of conceptual modularity: If the same individual responds in two different contexts to the linguistic input, "Smith is a bachelor" with the same output, "Smith is an unmarried man", then his linguistic behaviour is caused/explained by the fact that he has the same representation of the concept *bachelor* in both contexts. Importantly, here, mental states are individuated by linguistically represented concepts – epistemically possessing the concept *bachelor* consists in representing an *unmarried man* in a language-like manner. Subsequent to conceptual modularity, what does propositional modularity mean? According to Ramsey et al. (1991), propositional modularity says that propositional attitudes are "*functional discrete, semantically interpretable* states that play a *causal role* in the production of other mental propositional attitudes, and ultimately in the production of behaviour". They are functional discrete, because agents lose and gain individual beliefs; they are semantically interpretable, because people's behaviour can be generalized on the basis of folk psychology – specifying people's propositions (attitude and content) allows the lawful prediction of their behaviour; they play a causal role because specific propositions are supposed to explain specific behaviour causally.

What are the main problems for symbolicists? Firstly, it is largely ignored that cognition is a real-time process which always has a temporal dimension. Secondly, the high-level discreteness of symbolic computational models makes them quite brittle – that is, the destruction of individual representations collapses the whole system. (Eliasmith 2005, pp. 149-50) Thirdly, low-level perceptual processes are inadequately described: (i) Evidence for pervasive interactions between lower-level and higher-level types of cognition (e.g., the continuing multicortical area input to the thalamus and to motor structures) (Churchland et al. 1994, pp. 39-47) queries the classical conception of a mainly unidirectional, strict low-to-high processing hierarchy. (ii) Recognitions in real-world cases depend on rich recurrent patterns that involve visuomotor patterns (*ibid.*) (that is, affordances) (e.g., the chair with the slightly broken leg will break if I sit down). This is at odds with the assumption that the visual system is connected to the motor system only after the scene is fully recreated. (iii) If one thinks that one of the central functions of vision is to guide behaviour, then the assumption that the three-dimensional world is perfectly internally recreated might be problematic. To fulfil this function there is no need for a perfect recreation. Fourthly, if symbolicists invoke the classical hardware-software distinction and the irrelevance of neural implementation for understanding mental phenomena involved, then they are confronted with the fact that cognitive neuroscience explanations are multilevel in nature – *the* hardware level does not exist (PS Churchland 2002, p. 26). Without a correct view of the scientific practice of cognitive neuroscience, it seems to be inadvisable and unjustified to claim the explanatory irrelevance of neural processes. Fifthly, neither phenomenal nor non-phenomenal neural processes consist of formal symbol manipulation. The evidence for visual and motor imagery tells another story. All we have are sensorimotor vehicles and contents. Mapping propositional structure onto the brain to reconcile intentional realism and causal efficacy may not be considered the best solution. Embodied cognition warrants this compatibility better. Sixthly, even if it were phenomenally tenable to speak of formal symbol manipulation, it seems a bit adventurous to map those language-like structures onto the brain without any knowledge of neural processes. Seventhly, as we will dwell on more detailed later, there are no context-insensitive representations, and even less if they are conceived of as language-like. Hence, conceptual modularity is wrong. Eighthly, neither phenomenal nor non-phenomenal neural processes consist of tokens of propositional attitudes standing in relation to linguistic representations. All we have is auditory and visual processing of word forms and content bearing sensorimotor simulations. Hence, propositional modularity is wrong. Of course, that is not to say that we cannot give a reformulation of words like 'belief' or 'desire'

in terms of embodied cognition. Only the computational notion of propositions is abandoned. Ninthly, it is a fundamental mistake to limit causal processing to non-phenomenal neural processing. Not only because the problems of epiphenomenalism and causal overdetermination arise, but because the phenomenal experiences themselves belong to us as biological organisms which rely on them in order to coordinate behaviour. According to Craver (2009, p. 216), phenomenal experiences and non-phenomenal neural activation do not compete for causal efficacy. Then the question of mental causation that remains to be explained is how phenomenal processes are linked to motor areas and effectors. But that should not be too difficult, since all phenomenal experiences (including conceptualization) consist of sensorimotor activations.

While we have dealt with the older and more disembodied version of computationalism in this section, we are now going to be concerned with a kind of computationalism that is more bottom-up than symbolism, namely connectionism.

5.2 A critique of connectionism

I would now like to outline the basic features of connectionist modelling of mental processing.

At the centre of connectionist modelling of cognition is the notion of *parallel distributed* processing. It is contrasted to serial processing and forgoes explicit representations of rules and the compositional character of representations. As a kind of computation, connectionism aims to capture how cognition brings about responses/outputs formerly caused by stimuli. To this extent it adheres to the classical notion of information. Just as neurons comprise information about the firing of their inputs and transfer the processed information from their inputs to a wealth of other neurons, so too the computational unit receives inputs that either excite or inhibit its own activity, sums them up and transfer information about the sum via output connections to other computational units. Like information flows through different physically independent structures of the brain (retina – lateral geniculate body – optic radiation – visual cortex), connectionism models the information processing between independent layers of input, hidden, and outputs units. The input of the receiving unit is a product of the activation-level of the sending unit and the strength of the connection between the sending and receiving unit. Learning takes place if the strengths of the connections are changed.

Since knowledge is represented in a distributed way, representations are 'damage resistant' and 'fault tolerant'. In contrast to serial processing using local representations, memory can be assessed by content.

What are the weaknesses of connectionist models? Firstly, their disembodiment. The processing of inputs is often grasped in terms of purely linear or sigmoid response functions. However, real neurobiological activation patterns have heterogeneous, non-linear, spiking neurons. Secondly, methodologically connectionist models are explanatorily quite limited: Simple nodes are connected and trained in order to compute complex functions. Connecting ten billion nodes (the brain consists of so many neurons) and training them will get you nowhere. Thirdly, as the mechanistic framework strikingly illustrates, the comprehension of complex biological systems requires the interplay of bottom-up and top-down information. The individuation of conscious and behavioral phenomena and the individuation of neural networks are mutually dependent. Since connectionist models by no means consider the whole properties of cognitive systems qua behavioral or phenomenal ones – except for properties like statistical character, parallelism, or content assessable memory – and since they are bottom-up approaches, if at all, they disregard one essential component. (Eliasmith 2005, pp. 151-2) Fourthly, what is critical is that an adequate computational modelling of properties of cognitive processes (such as their statistical sensitivity, their fault tolerance or parallelism) is in no way the same as mechanistically explaining those properties. Clearly, one could say that connectionism, unlike CTM, should be understood as a model that can adequately model or simulate cognitive systems. It is not claimed that mental processes *are* connectionist, only that they are computable by means of connectionism. However, the question still arises as to why this should be sufficient. Fifthly, though it is the case that a simple feed-forward connectionist net realises productivity in a context-specific way, it is inherently disembodied. Why? The layer of input units is understood as a perceptual system that detects sensory features, or if you wish, constructs a phenomenal outlook. The layer of hidden units is regarded as giving conceptual interpretations of inputs. Decisively, the mapping from input to hidden units is arbitrary and merely mirrors the random weights that were assigned to the connections before learning. Insofar as connectionist models describe concepts as strings that bear an arbitrary relation to their perceptual and motor referents, concepts are disembodied. What these last remarks make indirectly clear is that connectionism itself can model almost everything – this means in reverse, however, that it itself provides hardly any substantial elucidations. Sixthly, the excitation-inhibition model is an oversimplification or a partial picture, given that 'modulatory' neurotransmitters change

the basic properties (e.g., the activation profiles) of the recipient neurons (Wheeler 2005, p. 87). Seventhly, also the point-to-point signalling model distinctive of mainstream connectionism does not do justice to the complexity of real brains, in view of the fact that diffusible modulators (e.g., the gas nitric oxide) demonstrate that neurons can interact and change one another's properties although they are not synaptically linked (Husbands et al. 2001).

In distinction from the first two sections of this chapter, we are now going to make a positive contribution. This consists in reconstructing the pivotal ideas of the dynamical systems theory of cognition.

5.3 Spivey's understanding of dynamical systems theory

Rather than the mind being composed of independent systems for perception, cognition, and action, the entire process is perhaps better conceived of as a continuous loop through perception-like processes, partially overlapping with cognition-like processes, and action-like processes, producing continuous changes in the environment, which in turn, continuously influence perception-like processes. (Spivey 2007, p. 10)

Spivey's *Continuity of Mind* (2007) provides an account of the dynamical systems theory of the mental abundant with information. Instead of focusing on computing static, discrete representations being temporally marked-off, the entire continuous trajectory which the mind undergoes is looked at. The relation between sensory stimulation and mental activity and the relation between mental activity and action are not thought of as linear, one-directional progressions that can be studied independently of each other. Rather, actions as temporally executing processes permanently change environmental stimuli (accompanied in parallel by proprioception) that again continuously alter mental activity, which is permanently manifesting and recasting its propensity for action. To put it simply, one perceives one's own actions when one is performing them – the actions brought about mentally continuously provide new stimuli for mental processing, which in turn changes behaviour. Sensory stimulation, mental activity and action cannot be taken to mean separable components in a linear sequence of stages, but have to be considered as participating in a continuous inseparable loop. I think that this inseparable loop absolutely applies to online processing.⁷⁴ What about offline cognition (e.g., pondering a philosophical problem)? (i) Clearly, the

⁷⁴ This concept of online processing is supposed to incorporate higher-level cognition. In terms of PSS, simulations permanently influence environmentally embedded processing. In respect thereof one could speak of online cognition. By the way, one has to thought of this ongoing influence of higher-level cognition as dynamic and temporally extended process (it is not a static mapping of a static simulator onto a static phenomenal representation).

contents of offline episodes are not triggered first and foremost by perceptions of behavioral input, but by memorized representations. In respect thereof they are not embedded in the respective environmental context. Hence, offline cognition does not fit into the depicted sensation-cognition-action loop. It does however make use of memorized sensorimotor contents and thus depends on the context of previous conceptual processing. (ii) Insofar as the brain consists of sensorimotor couplings that are inter alia reflected in phenomenal experiences, and higher-level cognition is nothing more than environmentally decoupled re-enactment of those sensorimotor couplings, we have an inseparable sensation-cognition-action loop. In respect thereof the three components are structurally inseparable. To put it succinctly, making use of sensorimotor areas and being environmentally embedded do not coincide.

Mental processing is a continuous trajectory through state space. This replaces the classical computational notion of static, context-insensitive amodal representations. Unlike connectionist modelling, the updating of neural activation is not considered as taking place in lockstep corresponding to arbitrary time spans. Rather, the dynamical systems theory presents cognition as a continuous trajectory through a state space grounded in real-time processing. Clearly, assuming something like a continuous trajectory is a fortiori at odds with the CTM that regards the mind as a central executive performing inferences on discrete language-like entities.

To get an intuitive understanding of the notion of *non-linear attraction* it is helpful to consider an example in visual perception, the Necker cube. Sometimes the cube is seen from slightly above, sometimes from slightly below. The two perspectives are constantly alternating. In terms of dynamical systems, this bi-stable pattern of perspectives corresponds to attractor basins that compete against one another. What is important is that though a classical logical symbol system can model flipping back and forth between quite stable states, it cannot pay attention to variations of intermediating transitions, as distinguished from dynamical systems. Additionally, it is indeed the case that the perspectives are not alternating with a step function (either the one perspective or the other), but rather the transition corresponds to a sigmoid function which involves descriptions of processing (autonomous periods) during the two perspectives. Since there is more than one attractor in a dynamical system at any time, it is grasped as non-linear (imagine a graph with at least two y-values at one moment (x-value)). This interpretation of non-linearity is also reflected in the parallelism by which mental occurrences are characterized: For example, phenomenal presentations and representations, non-phenomenal simulations of prospective sensory input, and phenomenal

simulations of past experiences often proceed simultaneously. Mental representations are “fundamentally continuous, graded, and partially overlapping (before overt action converts them into discrete actions)” (Spivey 2007, p. 10).

According to Spivey, 'pure mental states', that is, the maximized firing rate of a population of neurons for “I see a table”, for example, never take place. What de facto happens is that neurons are probabilistically activated. For instance, the time a listener needs to understand a word travelling to the respective attractor basin in his state space is greater than the time he has to understand that word without being influenced by a new word. In real-time circumstances, new stimulation is permanently in play – bottom-up sensory stimulation, stimulation brought about by mental imagery or by new motor behaviour. Considering that, the abundance of time spent in an attractor basin is probably far smaller than the abundance of time spent in covering distances toward attractor basins. To avoid misunderstandings, the mentioned 'pure mental states' are not irrelevant or inexistent. They are possible positions in a neural system's state space. As I said, however, they never really happen. To think it through to the end, if one considers real-time cognition, nowhere are states instantiated. What occurs are mediating mental processes across-the-board. That is not to say that brain states cannot be theoretically described – at a given point in time the neural activation across the brain corresponds to a brain state.⁷⁵ Within this framework, such brain states are presented as more or less approximated to pure mental states. Moreover, the notion of a temporally distinct stimulus that is waiting on its individual response also never takes places in the normal course of processing: “Instead, we continuously interact with a flowing train of multimodal perceptual arrays containing objects, agents, and events.” (ibid., p. 47) In contrast to the information-processing framework, whereby external stimuli encounter the senses in static copies which are subsequently processed through several mental stages and whereby every stage waits until the foregoing stage is finished before it can start its own work, mental processing is better grasped in terms of non-linear, continuous trajectories in mental state space.

Regard the following passage:

According to the continuity of mind thesis, that introspective impression of one discrete mental state after another is an illusion caused largely by the discreteness of the semantic labels we use in our internal monologue and by the discreteness of some of our-goal directed motor output. [...] Thus, although you might feel as though you think *p* and then *q* and then *r*, what you are actually thinking during that period is *mostly p* and *partly r*, and so on. (Spivey 2007, p. 308)

⁷⁵ For example, you are currently reading this sentence by means of auditory imagery while you are simultaneously undergoing phenomenal perceptions triggered by bottom-up stimuli.

Surely, as already mentioned (chapter two, section two), discrete linguistic forms used publicly purport the notion of static, discrete representations. Apart from this, one could readily doubt whether an attentive introspection really creates the illusion of step-by-step processing of discrete symbolic representation. I believe that it is the theoretical notion that mental processes are language-like which engenders such an illusion. By contrast, I think that phenomenally the continuity of mind thesis is highly plausible. Introspecting reveals quite saliently (at least in my case) that one continuously receives various stimuli arising from different sources nearly in parallel – like bottom-up sensation, memory, visual/motor/auditory imagery (e.g., inner speech), and so on.

According to Spivey, the trajectory of which the mind consists spans neural, bodily and environmental parameters. The state space that instantiates a mind comprises patterns of neural activation, muscular-skeletal kinematics, and objects in the environment. Just as sensory receptors and cortical neurons are causally related, so too qualitatively identical causal relationships exist between cortical neurons and muscles or between environmental objects and sensory receptors.

Spivey himself holds the following view: The centralised view of mind is substituted by a system of dynamical interaction between brain, body, and environment that generates control. The processor (vehicle) and the processed (content) are not independent of another insofar as external information is computed or transformed serially or in parallel in order to provide an interpretation of the world. Rather, the processing and the processed belong to the same state space of non-linear interacting patterns.

All that has been said so far is compatible with the assumption that our phenomenal being-in-the-world is internally generated by neural mechanisms. Again, this is not to say that neural mechanisms are not environmentally embedded. On the contrary, they are understood as constituting the whole mental system that is unconsciously and consciously directed to the non-phenomenally or phenomenally represented environment.

Asking whether mental content is determined by factors external or internal to the mind is not as adequate and straightforward as it seems at first sight. Taking into account that conscious episodes are intrinsically intentional, and therefore including the world as phenomenally presented, it is at least phenomenologically wrong to say that conscious processes are exclusively internally determined (Zahavi 2004, 2007, with Gallagher 2008). Conversely, insofar as external objects are not intelligible independent from our phenomenality, and insofar as they are always objects for us in order to be intelligible, it also seems inadequate to say that external objects are external in the sense that they are

independent of being represented in a certain way to human beings. That is not to say that it is meaningless to assume that there is a mind-independent world that triggers our phenomenal presentations and representations. Such a thought is intelligible, but it is intelligible only against the background of our phenomenal representations. From the fact that we conceive of mind-independent objects only in terms of our phenomenality, it does not follow that the thought of a mind-independent reality triggering our phenomenality itself is not intelligible. Thinking of a mind-independent world which triggers our phenomenal world model does not constitutively rely on grasping those entities, as they are independent from our phenomenality.

As we have seen, from the viewpoint of the phenomenality of the first-person perspective, a strict distinction between internal and external factors is inadequate. Obviously, there is a wide range of phenomena that are experienced particularly phenomenally as internal – the so-called ‘opaque’ episodes. Detached offline pondering or daydreaming while walking through a phenomenally represented environment are good examples for this.

What about the third-person perspective? Imagine that a neuroscientist observes the behaviour of a patient who is suffering from certain a pathology, namely, visual agnosia. Though the patient has no conscious visual experience of shapes and orientations of objects, he is nonetheless able to perform fluently motor actions directed at those objects. Thereupon the neuroscientist concludes that conscious and visuomotor vision are functionally dissociable, or that humans process visual information unconsciously. (cf. Milner & Goodale 1995) Most importantly for our purposes, the neuroscientist's observation of the patient's goal-directed behaviour functions in the frame of the neuroscientist's phenomenality – the objects the patient unconsciously acts on are objects only intelligible insofar as they are phenomenally represented for the neuroscientist. This can be called 'third-person phenomenality'. Moreover, taking a third-person phenomenality perspective allows us to speak of scientific entities not visible to the naked human eye. Although scientific ancillary apparatuses give access to entities otherwise not perceivable for us, the phenomenal representation and understanding of those entities are causally dependent on our human visual system.

The next chapter attempts to articulate the consequences of embodied mind for classical philosophical questions relating to mental content.

6 Implications of embodied cognition in general and PSS in particular for classical philosophical questions concerning mental content

If [...] we start with an empirically responsible philosophy – one that rests on the broadest convergent evidence – then the embodied and imaginative character of mind requires us to rethink the philosophy of language from the ground up. (Lakoff & Johnson 1999, p. 468)

6.1 An embodied theory of intentionality

In what follows, I will first outline some notions of intentionality belonging to the embodied mind paradigm and phenomenology. Then I will criticize classical theories of intentionality. Let us initially consider the phenomenal model of the intentionality relation (PMIR) that goes back to Metzinger (2003, 2004a, & Gallese 2003). What is a PMIR? Firstly, it depicts a relationship between a system and an object component. The relationship is depicted as currently held, and the system is always transparently represented to itself. 'Transparently' means that the system's self-model is not represented as an internal representation, but as if there were a self in the outer world. Secondly, human beings can represent the object component both transparently and opaquely. Opaque representations are such representations that are phenomenally represented as internal. Thirdly, by integrating the currently active, transparent self-model with an opaque action simulation (e.g., pondering offline on possible actions) and by coupling it to the effectors, a conscious, volitional first-person perspective (e.g., phenomenal experience of practical intentionality) emerges, that is, a neural simulation of a behavioural pattern configures the object component and this simulation is consciously represented as currently selected, in so far as depicting the practical intentionality relation. This integration of the simulated behaviour into the currently active bodily self-model leads to its functional and phenomenal embodiment, thereby becoming causally linked to the motor system and the effectors. The conscious experience of steady executive control can be representationally analysed as follows: The content of a transparent self-model is currently present in a transparent model of reality and thereby presently experiences itself as performing an action that it has previously simulated and selected opaquely, whereupon it becomes gradually assimilated to the subject component. Fourthly, a PMIR exhibits an experienced direction – arrows pointing from the self-model to the object component. Of course, in respect to complexity, this directedness is continuable, for instance, a first-order PMIR can take in the object component, whereby the second-order PMIR models a system-system relationship (that is, a system has a system-object relationship as its object). Fifthly,

different phenomenal models of the intentionality relation make different relations between subject and object globally available: For example, consider the difference between “I am someone who is currently visually *attending* to holding a pencil in my right hand” and “I am someone who is currently *understanding* the content of the sentence I am reading”. To sum up, at the phenomenal level, we can approximate something similar to propositional structure: A content specifier (the object), a person specifier (the transparent self-model), and an attitude specifier (the phenomenally represented relation between subject and object).

I construe the PMIR as a phenomenal relation that is intrinsically meaningful. This model can be processed in the form of non-phenomenal and phenomenal representations or simulations. The PMIR should not be confused with my idea of internally referring linguistic expressions (that is, internal associations between arbitrary tokens and content bearing sensorimotor simulations).⁷⁶ Actually, concrete instantiations of the PMIR have to be taken as the referents of corresponding linguistic expressions.

Akinetic mutism (anterior damage to cingulate gyrus or bilateral medial parietal damage) is a disorder which expresses itself at the behavioral level in the form of glances without any appearance of intention. Admittedly, patients suffering from this disorder can momentarily track objects or say their name, although they are not volitional subjects which are able to exert control. Those patients have an integrated functional self-representation that makes self-related information globally available for the control of action or guided attention. In Damasio’s terms (1999), they exhibit wakefulness, but not ‘core consciousness’. According to Metzinger, core consciousness is the basic PMIR. Thus what those patients lack is a phenomenal representation of a self being autonomously related to possible action aims (in other words, there is no volitional subject). They fail to integrate opaque simulations of possible actions into the transparent self-model, for example.

I will now consider the thesis that intentionality should be phenomenalized. What is most important is the way intentionality is *itself* represented on the level of phenomenal content. And in respect thereof, intentionality can be naturalized, so the claim. In other words, the phenomenal experience as a subject with re-presenting perceptions, thoughts or intentions can be naturalized.

Firstly, the brain depicts the relation between subject and object as an asymmetric one. We experience ourselves as if we project visual attention outwards or inwards. The brain creates a self-model that causes the appearance of cognitive agents which, at will, epistemically

⁷⁶ In other words, the reference relation of the PMIR is not analogous to the reference relation of language and content.

focus on the external world or on internal lives. For that reason we, as human beings, find the idea of an objective intentionality relation intuitively plausible. Secondly, similarly hypothetical, philosophical debates on intentionality have been influenced by our predominant sensory modality, namely, phenomenal vision. Just as our transparent visual model of reality generates distal objects, so too, the representational relationship is internally modelled. Subsequently, if the object component is opaquely represented (such as in offline simulations), a philosophical interpretation of these mental contents as non-physical (“intentional inexistent”) would seem natural.

What are the neural underpinnings of the phenomenal representation of the intentionality relation itself? In the case of F5 canonical grasping neurons (which respond to visual representations of objects' size and shape without any movement), the selectivity for an executed grip and the selectivity for the visual representation of objects are, to a considerable degree, congruent. To this extent, functionally, the F5 canonical grasping neurons are not responsible for either sensory input or movement, but should instead be considered as relational. Furthermore, physical entities are not coded in respect of their mere physical appearance, but in terms of interaction with an acting agent.⁷⁷ Provocatively, one could claim that classical philosophical theorizing about intentionality (including the problems of intensional contexts and non-synonymous co-extensional expressions) is a consequence of a naively realistic understanding of the process of visual attention, of the phenomenal self-focusing on a phenomenal visual object.⁷⁸ (Metzinger 2003, pp. 411-21)

What types of directedness are there? Firstly, mental and phenomenal experiences in the sense of the PMIR (the transparent or opaque representation of the object is that at which the transparent self-representation is directed). Secondly, to the extent that the brain generates presentations and representations by responding to external stimuli it may be considered as being directed toward the stimuli. Thirdly, an example of directed non-phenomenal occurrences are unexecuted motor programmes which the brain automatically provides if it perceives respective objects (the two functional clusters FA-VIP and F5ab-AIP) (this could be dubbed 'motor intentionality'). Another example of hardly consciously experienced directed mental occurrences are cases in which phenomenal representations are fused with conceptual ones, such as expecting your mobile phone to buzz and thereby actually experiencing its buzzing, even though it did not buzz.⁷⁹ An instance of fairly consciously experienced directed mental processes could be turning a phenomenal simulation of a racing

⁷⁷ Delacour (1997, p. 138) describes the notion of neural structures which model the intentionality relation itself.

⁷⁸ It looks as if the indeed great epistemological problem of relating mind and world would be confined to explaining how language-like concepts can be adequately applied to the objects of our phenomenal world model.

bicycle, for example, towards a phenomenal representation of a token of this type of racing bicycle – that is, a highly conscious recognition process. This is readily compatible with the possibility that in this case simulators are also likely to be brought into consciousness by automatic, subpersonal processes.

What sorts of derived intentionality there are? Linguistic symbols only have content against the backdrop of concretely occurring mental processes. They are only semantically processed if they are associated with sensorimotor simulations. How do we actually understand non-phenomenally experienced kinds of intentionality, such as blindsight? I think we have to conceptualize such phenomena in terms of phenomenal structures (e.g., in the form of an arrow from a phenomenal model of a human body to a phenomenally familiar object), since intelligibility amounts to nothing more than this. This theoretical model could utterly fail, if one bears in mind that the stable structures of our phenomenal world model (single objects that we experience as identical in multimodal terms) are neurally constructed in the first place. In terms of theoretical intelligibility, understanding non-phenomenal intentionality is derived from sensorimotor structures of our phenomenal world model. Which is basically what Lakoff and Johnson have in mind.

In order to provide indirect reason for our theory of intentionality, I am now going to criticize other theories of intentionality.

6.2 Criticizing some classical theories of intentionality

(1) At first, we have a look at Putnam's Twin Earth case and Kripke's theory of referring proper names. What is problematic with this view?⁸⁰

It seems inconsistent to me that the beliefs of two individuals whose contents are phenomenally processed could have different contents although they are psychophysically identical. Why and how should it make a difference to both non-phenomenally and

⁷⁹ Given that this fusion constitutes the phenomenal perception, it seem problematic to speak of concepts being directed toward phenomenal representations. One could say that the more the phenomenal representations are constituted by concepts, the more inappropriate is the talk of 'directed'.

⁸⁰ It is important to emphasize that my argumentation presupposes the following way of thinking: Whether this thought experiment shows that natural kind externalism is true or not depends on the theoretical assumptions being made in the background. Because I deem the concept of embodied simulation right (in light of empirical evidence), I will clarify the implications of the thought experiment against the backdrop of the concept of embodied simulation. Moreover, as we shall see later, the more or less shared intuition that the thought experiment speaks for natural kind externalism that is almost devoid of theory cannot be considered as an empirically reliable source. If thought experiments, however, are based on theoretical assumptions, then they have to be pitted against empirical evidence. If they are not, then one has to appeal to the theory that is best confirmed. This way of thinking holds true for all my reflections about thought experiments and criticism at proponents of classical semantics. These remarks link to my understanding of philosophy that was delineated at the beginning of the Introduction.

phenomenally processed content that experientially identical objects (water) have a different chemical microstructure (water as H₂O and as XYZ), which again cannot be reflected in different psychophysical activation? What makes a certain sense is to assume that different physical structures can bring about the mental content by interacting with two cognitive systems (Metzinger 2009, p. 20). Then, however, the two cognitive systems internally process the same content due to identical psychophysical structures triggered by different external entities. Is it not somewhat strange to postulate a causal difference that extends to the content of beliefs? Why not? Given that the brain re-presents in a way that serves as a basis for the organism's behaving in an environment, the question that arises is how could it make any difference whether Oscar thinks about washing one's hands with water while he is doing it with H₂O, or whether Twin-Oscar thinks about washings one's hands with water while he is doing it with XYZ? I dare say that the assumption of a such a difference is motivated by feeling oneself committed to a realism and intersubjective objectivism of cognitive mental content. Moreover, because it is reasonable to ground conceptual thought, to which beliefs clearly belong, in phenomenal experiences and to regard actualised (that is, not merely dispositional) beliefs as being phenomenal⁸¹, it cannot be true that Oscar and Twin-Oscar have different beliefs.

Kripke's theory of the reference of proper names is also questionable. Internally, how could it be in any way relevant with respect to your understanding of the proper name "Gödel" that Gödel once was christened "Gödel"? Presumably you have seen a framed picture of him somewhere with some written information that you kept in mind in combination with the auditory and visual symbol "Gödel", and all this allows you to go through contentful mental episodes about Gödel. Referring to a causal chain that began with the christening in order to explain the content of mental processes shows very plainly how cognitive and phenomenal processes were entirely neglected.

(2) Now, I am going to criticize the classical informational semantics that is characterized by the following: X (an object, event) is the intentional content of C (in other words, a state C carries information about X) iff (i) Xs covary nomologically with tokens of C, (ii) C has the function of carrying information about X (Dretske 1995), (iii) if anything else causes C, its doing so is asymmetrically dependent on X's doing so (Fodor 1990), and (iv) an X was the incipient cause of C (Prinz 2002).

What is wrong with this conception? Firstly, the theory cannot explain how we learn new information by reading or communicating, since the mental episodes process intentional

⁸¹ This is believed to be true not only of our understanding of embodiment, but of all theories of beliefs.

content in the absence of co-varying environmental objects. Reacting to this by saying that the notion of causal co-variance also involves the co-varying of internal mental occurrences may solve this problem, but we still lack concrete internal mechanisms. The notion of causal co-variance is quite simply too wide and therefore uninformative. Secondly, causal co-variance cannot explain our capacities of forming contentful mental episodes about scientifically impossible, fictive, future, or ideal objects. It is simply not possible, if one bears mind that our mental episodes cannot covary with such objects, since they do not exist. Both objections cry out for internal mechanisms that can generate simulations which, inter alia, exhibit productivity (creating non-experienced objects or events) in the offline modus – precisely that is what PSS provides. Thirdly, it seems natural to think that the model of causal co-variance is intellectually motivated by conceptually representing a co-variance between objects or events of our phenomenal world model and our phenomenal self-model. If this were the case, it would take the objects and events of our phenomenal self-model as mind-independent objects that are actually internally generated. The supposedly co-varying stimuli are internally generated. Fourthly, even in phenomenal terms of the relation between these two internally constructed models, it is simply wrong that our opaquely represented mental episodes are always correlated with phenomenally transparent represented objects or events (e.g., pondering on a philosophical problem while sitting in the canteen). The parallel distributed character of the mind is plainly too complex to be able to be grasped in terms of simple object correlations. Fifthly, whether causal co-variance is an adequate metaphysical model for the relation between mind-independent entities and mental episodes is epistemically beyond reach, since we would need a grasp of inconceivable entities. That is not to say that we cannot develop scientific models describing how physical entities (invisible to the naked eye) are related to the objects of our phenomenal world model. Sixthly, how can the co-varying entities be causally efficient for your concept, if you had not focused your selective attention on that aspect of your visual field which causally co-varied firstly with you as cognitive system? Even if you had focused your selective attention on it, the thereby extracted content permanently alters within cognitive systems – both caused by further phenomenal external stimuli and by merely internally produced alterations (e.g., fading memory). Thus the notion of causal co-variance does not suffice. Seventhly, why should one assume something like asymmetrical dependence, if one does not bind oneself to a realism assumption of cognition? Could it not be that the representational content of classifying an object as a COW that under normal conditions is classified by humans as a HORSE by mapping the internal simulator of a cow to the current phenomenal visual

representation, is simply the simulation of the cow (that penetrates the phenomenal visual perception)? (Or maybe you undergo a blurry presentation on which you have no concepts, and this is what your mental content amounts to.) It seems appropriate to say that content which is cognitively available in this respective context is the COW representation, whereas the content not cognitively available for the respective subject is the HORSE representation. Eighthly, in the case of internally triggered simulations, causal co-variance between external entities and internal processing is simply inadequate. Even if phenomenal representations causally covary with mind-independent objects (in the case of ongoing online processing), environmentally decoupled offline simulations only make use of internally memorized representations. Ninthly, I think it is important and perhaps illuminating to ask for an explanation of why we find causal co-variance intuitively compelling, if indeed we do. In the first place it is a consequence of our internally modelled PMIR. This model obviously also functions in the context of theoretical considerations. The notion of causing external entities of classical informational semantics is possibly not only not intelligible independent of our phenomenal representations, but is itself generated by our phenomenal experience of acting in an external world. Of course, emphasizing the importance of intelligibility concerning external entities does not preclude that some manner of external entities exist. It therefore also does not preclude that intelligibility itself in the form of sensorimotor structures goes back to an inaccessible relation between cognitive systems and an objective external world. But how could that be tested, if not by making use of phenomenal mental models? Tenthly, postulating incipient causes or asymmetric dependences amounts to nothing more than arbitrary ad hoc assumptions which are supposed to warrant epistemic realism. Does that mean that if I represented my first encounter with horses as an encounter with cows, then my mental occurrences had no content at all? This seems to be unreasonable. Furthermore, how is it possible that if I did not pay selective attention to the phenomenal visual representation of my first encounter with a horse, that I nonetheless extract the concept of a cow? In other words, how can incipient causes or concepts applied under normal conditions be thought to be causally efficient, if the subject does not cognitively process those items?

Besides, if one subscribes to the idea that our offline mental episodes are re-enactments of sensorimotor areas whose contents are to a large extent determined by previous phenomenal experiences, as I do, and understands the notion of causal co-variation in respect thereof, then the thought of causal co-variance is indeed useful. Furthermore, if one adheres to a broader concept of causal co-variance that also includes the co-variation between internal representations, then the co-varying between diverse modality-specific occurrences (e.g., a

visually represented philosophical position and a certain emotional state, such as esteem) may be considered as an instance of this concept.

(3) Now, consider conceptual role theories (Harman 1982; Loar 1981): The meaning of a term is determined by the inferential relations it is integrated into.

As long as it is not clarified *how* inferential relations are internally processed in order to be contentful such an account leaves too many questions unanswered. And if they are defined in terms of a computational notion of propositions, then, as already seen, the theory is inadequate.

(4) Two factor theories say the following (Field 1977; Block 1986): Causal relations and conceptual roles together determine the meaning of a term.

Two factor theories are faced with the already depicted problems of causal accounts *and* conceptual role theories.

Now let us shift attention to the debate on whether a narrow or a wide account of mental content is right.

6.3 Narrow mental content

Since we have repudiated Putnam/Kripke's notion of reference that is implied by natural kind externalism and since we have eliminated reference to external entities (be it mind-dependent or mind-independent ones) in the case of environmentally detached sensorimotor simulations and because we have also argued for the internality of phenomenal representations and the grounding of conceptual structures on them, it is obvious that we hold a narrow view of mental content.

In the following section I am going to first discuss why PSS should be considered as a theory that fixes simulational content narrowly. I thereafter present Burge's thought experiments for social externalism about mental content. Then I will outline two vital objections to semantic externalism. We are concerned mainly with conceptual content.

Phenomenal presentational and representational content is narrow content to the extent that it necessarily depends on a virtual presence generated by the brain. The mapping of simulators onto a stable phenomenal world model is narrow content in that it is based on sensorimotor stand-ins memorized in the respective cognitive system.

Since offline simulations are re-enactments of sensorimotor areas belonging to specific individuals which have different experiential material that they can re-enact (or re-combine), and concrete contextually embedded simulations specify mental content, mental content has

to be individuated narrowly. The complex multimodal bodily experiences that can be re-enacted are internal processes.

Let me briefly describe the thought experiment intended to underpin *social externalism* as conceptualized by Tyler Burge (1979, 1986): Imagine an English-speaking individual called Jane who utters the sentence, “I have arthritis in my thigh.” Since arthritis is a state of the joints only, her utterance expresses a false belief. Now imagine a counterfactual situation where Jane has the same internal state and personal history, yet her surrounding community uses “arthritis” for a different ailment that is called “tharthritis”. This disease comprises rheumatoid afflictions of both thighs and joints. In this context, her above-mentioned utterance would be true. I am referring to this because it is supposed to show that the contents of her beliefs differ in the two contexts, though she is psychophysically identical. Since the only thing that differs in the two contexts are the linguistic usage and the content of her belief, it is inferred that mental content is partly dependent on shared linguistic practices. This, again, is to demonstrate that semantic externalism is right.

What pivotal strategies do the two objections respectively pursue? The first one disagrees with the intuition that the contents of the beliefs differ in both worlds – it is argued that they are actually identical. The non-phenomenal and more or less phenomenal processing accompanying Jane's utterance, “I have arthritis in my thigh”, is identical in both contexts. Individuating her phenomenal process could look as follows: Imagine that her context-specific sensorimotor simulation consists of an auditory and visual simulation of the word “arthritis” that is associated with a feeling of pain, a proprioceptive localization of this pain in her thigh, and a dynamically processed body image. If one takes for granted that this highly context-specific phenomenal processing is identical in both contexts, and that what one calls 'belief' equates to this processing, then it is simply wrong that her psychological content differs in these two contexts. It is irrelevant whether “arthritis” is conventionally used for joints only or for both thighs and joints.

The second objection concedes that it is right anyhow that her statements, which make use of linguistic forms whose sensorimotor meanings are determined socially, have different truth values in both contexts – the former is wrong, the latter right. Thus if the sensorimotor truth-condition of “arthritis” is socially determined in terms of joints only, then Jane's statement, “I have arthritis in my thigh”, is wrong in this social context. To put it clearly, which linguistic forms are correctly associated with which sensorimotor simulations is the result of

social processes. However, that in no way implies that the mental contents which precede her statements are wide.⁸²

I think that this place is suited for carrying out the announcement from the last section of Chapter one – that is, arguing against extended functionalism – because this position is also a strong form of externalism. The line of argument is as follows: Because both Inga and Otto⁸³ have to make use of inner sensorimotor simulations in order to make the respective information (stored in memory or contained in a notebook) intelligible, and in view of the fact that one could specify neural processes at a certain level so that they uniquely realise phenomenal experiences (e.g., in terms of modality-specific feature maps), then we have no different realisers of the same functional state. Hence, if cognitive processing presupposes intelligibility, and intelligibility is tied to sensorimotor simulation in the case of environmentally detached mental processing, then it is neither the case that the respective information is multiple realizable nor that cognitive processing extends into the environment. In my view this example only shows that a kind of embedded interactionism seems to be true – that is to say that phenomenally transparent representations of linguistic symbols control cognitive processing (see 4.5).

6.4 Nonconceptual mental content

The following section is supposed to provide a theory of nonconceptual mental content. For that we refer to some distinctions introduced in the second Chapter and in the context of the reconstruction of PSS.

Determining which mental contents are nonconceptual depends on the notion of concepts one adheres to. Thus the notion of nonconceptual content is an essentially 'contrastive' (Bermúdez 2008) one.

Remember that according to PSS, in high-level functional terms a concept is a simulator – that is, a frame which integrates perceptual symbols across category instances and the dispositional property of generating an infinite set of concrete simulations. A necessary and sufficient condition for being perceptual symbols is their being extracted by means of selective attention and their permanent storage in long-term memory.

⁸² Burge's utilization of the thought experiment relies on the adequacy of a usage theory of meaning (e.g., Sellars 1954). He has no concept of a belief apart from an empirically highly questionable folk psychology. We will see later that it is quite meaningful to speak of non-linguistic mental occurrences having truth values.

⁸³ This example is completely reconstructed in Clark and Chalmers' essay "The Extended Mind" (1998).

Thus any internally processed content that is not stored in long-term memory is nonconceptual content. Of course this includes phenomenally processed content that admittedly is attentionally available, but that is not actually being paid selective attention to. All internally processed content that is not available for cognitive processing is nonconceptual content. As already mentioned, only phenomenally processed content is cognitively available ('principle of phenomenal reference'). Hence, unsurprisingly, exclusively non-phenomenally processed content constitutes a part of nonconceptual content. Moreover, to a certain degree, the internal dynamics of the brain is attentionally unavailable ('autoepistemic closure' (cf. Metzinger 2003 & Northoff 2004)).

In summary: Nonconceptual contents are all those contents that are attentionally unavailable, attentionally available but not actually being introspected, and actually being introspected but in which case we are unable to store them in long-term memory. Phenomenally experienced emotions that do not rely on the involvement of the neocortex (LeDoux 1996) may be considered as an instance of the second kind of nonconceptual contents.

Certainly the most interesting question is to grasp phenomenally processed nonconceptual content. What are examples of those contents? Phenomenal tokens like green₂₄ or green₂₆ of which we do not possess concepts (in the sense of temporarily stable psychological types that provide transtemporal identification). Although we probably can discriminate between those phenomenal tokens, we do not have concepts of them. Such contents satisfy two of the three functional properties of consciousness – they are attentionally available and available for motor behaviour (e.g., speech production or pointing movements) in discrimination tasks, but not available for cognitive processing. Such contents are not available for long-term memory and hence nor for cognition in general ('memory constraint').⁸⁴

Let me now present a special case of nonconceptual content for the sake of completeness: There is evidence that individuals with blindsight react emotionally to faces presented in their blind field even though they have no phenomenal visual perception of the faces (de Gelder et al. 1999).⁸⁵ If those individuals have phenomenal experiences of their emotions and if those emotions were extracted by means of selective attention, then only the stimuli (to be

⁸⁴ This is more or less in alignment with the classical argument that the content of perception is more fine-grained than the content of propositional attitudes (e.g., Peacocke 1992, Tye 1995, Heck 2000). McDowell's proposal (1994) that demonstrative concepts (such as *that shade*) allow us to conceptually represent colours with the same fineness of grain with which they are perceptually represented does not work. The reason is that internalizing the linguistic behaviour of saying *that shade* is not sufficient for the storage of concepts of the respective phenomenal contents. Since concepts have to satisfy the 're-identification condition' (Kelly 2001), that is, they must be applicable in thought in the absence of the respective phenomenal sample in order to count as concepts, demonstratively captured presentational shades are nonconceptual.

⁸⁵ Clearly, it is assumed that those individuals have no non-visual phenomenal experiences (e.g., somatosensory) of the faces.

understood as mentally presented or represented objects of directedness (see 6.1)) from which the individuals have no phenomenal experience are nonconceptual contents. If those individuals have no or little phenomenal experience of their emotions, then both the stimuli and the emotions may be considered as nonconceptual contents. (Prinz 2007, p. 57)

There remain two further functionally individuated kinds of nonconceptual contents – on the one hand, those that are only available for motor control, and on the other, those that are only attentionally available. A good example of the former are the flexible and selective behavioral reactions a table-tennis player has to have in order to return a ball successfully. Examples of nonconceptual contents which are only available for attention are very subtle alterations in proprioceptive experiences or in cases of meditation (states of almost maximum focusing on phenomenal experiencing).

We can now state more precisely what it means to speak of being more or less conscious. The more functional constraints are satisfied, the more conscious those contents. The lower the degree of satisfaction of the functional constraints, the simpler the contents. Clearly, these are not the only constraints on which the degree of consciousness depends – even if only one constraint is satisfied, this phenomenal experience may prove to be strongly conscious if it is experienced as present to a high degree. (Metzinger 2003; pp. 30-43, 62-83)

The following remarks are limited to phenomenal nonconceptual content. Heck (2000) distinguishes between states⁸⁶ that are (i) state-nonconceptual and those that are (ii) content-nonconceptual. (i) Undergoing state-nonconceptual states does not require the possession of the concepts involved in a correct determination of the contents of those states. Conversely, state-conceptual states are those states the organism cannot undergo without possessing the concepts required for the specification of the state's contents (e.g., environmentally detached offline thought). (ii) Undergoing content-nonconceptual states means that the experienced content is nonconceptual. For a state to be content-conceptual means that its content is constituted by concepts. Are the previous remarks on nonconceptual contents cases of state- or content-nonconceptualism? Or is it a useful distinction at all? Bermúdez (2007) argues that the problem with the state-content distinction is that concept-dependent and concept-independent states can only be distinguished in terms of different types of contents. I don't think so. Concept-dependent and concept-independent states can be functionally distinguished – that is, if one has the capacity (the function) to undergo mental episodes whose contents do not come from environmentally embedded inputs, then these mental

⁸⁶ Because it is a reconstruction of Heck's position, I also speak of 'states' instead of 'processes' or 'episodes', whereas I think that the meanings of the latter expressions are neurophenomenologically more adequate in order to grasp the mind.

episodes are concept-dependent states in a manner that has nothing to do with different types of contents. Examples of state-nonconceptual states are non-phenomenal proprioception or vision-for-action. In terms of our terminology, the distinction between (i) states and (ii) contents is analogous to the distinction between the following two questions: (i) Is the globally (un)available content constituted by concepts? (ii) Is this content cognitively available or cognitively unavailable? As mentioned previously, environmentally detached offline simulations are state-conceptual. The same is true of phenomenal representations that are cognitively penetrated. State-nonconceptual are all low-level online processes. Content-conceptual processes are those whose contents are cognitively available. Content-nonconceptual processes are those whose contents are cognitively unavailable (e.g., phenomenal presentations).

How can content-nonconceptual contents be individuated? With reference to the example of the colour shades green₂₄ and green₂₆, one needs neurophenomenological correlates.⁸⁷ In the case of contents viewed phenomenally as internal (e.g., fleeting proprioception), one needs neurophenomenological explanations. What seems to be problematic is that repeating phenomenological reports can transform once cognitively unavailable content into cognitively available content.

Can state-nonconceptual content be completely autonomous of conceptual content? I should think so. If a creature lacks any memory capacities necessary for undergoing offline simulations, but nevertheless processes contents which are relevant for the coordination of its behaviour, its state-nonconceptual content is autonomous of conceptual content. Even if creatures can only aptly be assigned spatial contents if they are able to re-identify particular locations of the represented environment, and even if this requires that they can represent the changing of their positions within that environment, both representations can function at the nonconceptual level. What might be interesting is to elucidate modality-specific notions of state- and content-nonconceptual content and to explain how and why different modality-specific contents are comparatively more or less cognitively unavailable.⁸⁸

In the next section we are concerned with the question whether perception is cognitively penetrable.

⁸⁷ Of course, the specification of the phenomenal side has to be made by somebody who is experienced in individuating phenomenal experiences. Moreover, one could consult a colour scale in order to specify fine-grained phenomenal colour gradations.

⁸⁸ It seems to be the case that the content provided by phenomenal visual perception makes up the largest part of our cognitively available content.

6.5 Phenomenal perception is cognitively penetrable

Evidence for strong interactions between lower-level and higher-level neural areas (cf. Churchland et al. 1994, pp. 39-47) suggest that phenomenal visual perception is cognitively penetrable. If it is indeed the case that phenomenal visual perception is cognitively penetrable, then our central assumption of a single⁸⁹ representational format of perception and conception would be quite plausible, if only because the substantial additional cost of permanently transducing one representational format into the other were not required.

First of all, a simulation is a top-down activation of feature maps to re-enact perceptual experience. Besides top-down initiation, cognition affects the content of vision: Neuroscience investigations of visual and motor imagery demonstrate that in the absence of physical input cognition builds content in sensory-motor areas. According to Barsalou, the Müller-Lyer illusion only shows that bottom-up dominates top-down information when they conflict (aside from hallucinations), the illusion does not indicate the impenetrability of perception. In the absence of bottom-up influence, top-down information dominates (in the case of imagery); if bottom-up and top-down information is reconcilable, top-information penetrates, but rather complementary to bottom-up influence. Located between pure bottom-up and pure top-down efficacy, consider a cognitive process that relies on mixtures of bottom-up and top-down information to construct perception: filling-ins. In filling-ins, gaps of bottom-up information are filled by information from perceptual memory, typically an instance of cognition. For example, phoneme restoration adjusts low-level feature detectors – thus, filling in missing phonemes in word recognition, based on memory representations, not only penetrates conscious experience, but also sensory processing itself. Such findings indicate that cognition and perception are part of a common representational system and that they become merged in order to construct perceptual representations. (Barsalou 1999, pp. 588-9)

On a related note, perception's penetrability might be underpinned by considering the tilt aftereffect. It was shown that the illusion was significantly greater on the side to which participants paid attention than on the unattended side. Though the voluntary control of spatially endogenous attention is not equatable to higher-level knowledge, it begins within the association cortex. Critically, since the aftereffect illusion occurs in the visual cortical area, it is obviously influenced by feedback signals from the association cortex (e.g., Lamme and Roelfsema 2000). That is not to say that the feedback signals have the capability of fully

⁸⁹ With respect to the assumption that both representational formats are spatial in nature, one could speak of a *single* representational format. It would be doubtful to say that phenomenal visual perception and higher-level cognition apply the *same spatial* representational format, because the brain has different spatial formats available (cf. Paillard 1991).

replacing the local informational pattern in the visual cortex, but that they are able to subtly alter it.

In the next section we are concerned with the question whether the indeterminateness of conception is a conclusive objection to PSS.

6.6 Indeterminateness of conceptual processing

For we are opposing a language-like view of the mental vigorously, and the indeterminateness of conceptual processing was often cited as evidence against a visual and for a language-like character of conception, we have good reason to show that a sensorimotor, non-linguistic notion of conceptual processing permits indeterminateness.

It is often argued that since perceptual representations are picture-like, they are determinate. However, because human conceptualization is indeterminate, it cannot be based on perceptual representations. Human conceptualization is as indeterminate as our perceptual experiences are phenomenally richer than our conceptual capacities.

At first, let me introduce some important distinctions. It is one thing to question whether phenomenal episodes themselves are determinate, it is quite another to question whether phenomenally processed content is cognitively available in a determinate way, such as countability. In other words, the concern is whether phenomenal contents involve cognitively recognizable contents (see 2.2). Additionally, it appears to make a difference whether one attends to phenomenal presentations and representations or whether one attends to, if it makes sense at all (because one could also say that a certain kind of attention brings phenomenal simulations forth), more or less phenomenal simulations.

Perceptual symbols as constituents of concepts can of course be indeterminate for a variety of reasons: Schematically extracting stripes from a perception of a tiger may result in patches, as distinct from a representation of the exact number of stripes. Representing the tiger later, the free-floating symbol may depict the striped-ness of the tiger, without representing a particular number of stripes. If this symbol of being striped is attached to a surface of a simulated tiger, it would have a determinate number of stripes, however, probably inaccurately in respect to the original perception.

Now, consider an even more convincing solution, not assuming conscious representation: High-level neurons in perceptual systems can code information qualitatively – using the example, perceiving the tiger causes the firing of detectors which respond solely to it being striped, without capturing any specific number of stripes. Equally important is that

qualitatively coding neurons also offer an attractive solution for generality. Imagine the representation of a rectangle, whereby specific neurons code the representation of the lines independently of their length, position, and orientation. Furthermore, imagine that other specific neurons represent the vertices that join the lines, independently of the angles that link them. Hence, qualitative coding enables both indeterminacy and generality. (Barsalou 1999, pp. 584-5) To summarize: These points show the indeterminacy of sensorimotorily embodied concepts. From this one can also conclude that, as the notion of selectively attending to phenomenal simulations makes sense, the content of phenomenal simulations does not have to be cognitively available in a determinate manner.

In addition, consider the following three points: (i) Phenomena like change blindness (for example, Rensink et al. 2000) and inattention blindness (Mack & Rock 1998) make plain that people do not attentively experience or monitor presented details. What is remarkable in the case of change blindness is that we often fail to notice changes even when we are looking at the changes (tested by eye trackers) as they appear. (ii) Consider also the *virtual* character of phenomenal perceptual contents. Strictly speaking, it is not part of your conscious experience of a portemonnaie laying on table, for example, that it has a back side. The virtual presence of its back is a function of our implicit sensorimotor knowledge that consists in the ongoing activity of simulations, at least in terms of embodied cognition. (iii) If you reflect on the fixed content of your conscious episode which is currently occurring, then do you encounter nothing but determinate detailedness? I don't think so. Aspects of the background or of the periphery of your visual field are indeterminately present. Do you have a phenomenally clear grasp of the colour of the object that is shimmering on the outermost right side of your periphery? Do you have a sharp image of the whole passage of a book page if you reflect on what is determinately present as you attend to a sentence? I believe not.

Clearly, even blurred phenomenal visual episodes are always determinate. I think that all that change and inattention blindness, the virtual character and blurring on the periphery show is that phenomenal visual episodes are not attentionally and cognitively available in a determinate manner. They do not show that phenomenal episodes themselves are indeterminate.

If the notion of attending to more or less phenomenal offline simulations makes sense, and to the degree that some aspects are blurred while others are schematically represented, the latter would be cognitively available. If such a notion does not make sense, and one's attention is not particularly concentrated on the simulations, but the simulation itself is the cognitive processing, then, though the simulation is definitely determinate, only schematic information

is processed and so the cognitive system can exhibit indeterminacy. Furthermore, were it the case that every offline simulation (e.g., while reading a sentence or while quasi-seeing the back of the portemonnaie) produced a perceptual episode fully rich in phenomenal visual detailedness (non-schematical), you would permanently undergo hallucinations; and that would not be at all useful to you as an acting being.

As we have broached the objectivist cognitivist semantics in the first section of chapter one, we are now going to criticize the explanations of a possible world semantics.

6.7 Criticizing the explanations a possible world semantics provides for non-synonymous co-extensional expressions and intensional contexts of beliefs

(i) Calling to mind one of the standard examples of co-extensional but non-synonymous expressions: Although “renate” and “cordate” apply to exactly the same range of referents, the former means “creature with kidneys” and the latter “creature with a heart” (it is at least taken for granted that these two expressions are co-extensional).

A possible world semantic explains the different meanings in terms of diverging extensions in possible worlds – that is, though “renate” and “cordate” apply to the same things in the actual world, their referents differ in at least one possible world. (cf. Lycan 1999)

(ii) Reconsidering the philosophical classic of how belief states generate intensional contexts:
(1) The sentence, “John believes that the football player Kaka changed from AC to Real Madrid in 2009”, is true. (2) The sentence, “John believes that the Brazilian football player who scored seventy Serie A goals and twenty-three Champions League goals for AC Milan, changed from AC Milan to Real Madrid”, is false. Either because John does not have this belief or because he thinks something that contradicts the propositional content. Although the propositional contents of both sentences refer to the same individual (co-extensional), namely Kaka, only the first is true.

A possible world semantics explains the fact that co-referential sentences have different truth-values by saying that the linguistic expressions “Kaka” and “the Brazilian football player who scored seventy Serie A goals and twenty-three Champions League goals for AC Milan” indeed co-refer in the actual world, but not in at least one possible world – which is to say that their intensions differ. Since believing is a relation between a believer and a sentence’s intension/proposition, then it is quite possible that somebody believes the one, but not the other proposition.

What do these explanations suffer from? First, let me briefly outline already mentioned objections that are also true of possible world semantics: Appealing to actual or possible external referents in the case of offline simulations makes no sense; objectivist motivations for assuming possible world semantics; neglecting the requirement of internal processes constitutive of all kinds of mental content. Secondly, endowing quasi-linguistic mental processes with content exclusively by relating them to sets of referents signifies the static and context-insensitive character of classical accounts of mental content. Thirdly, explaining the two puzzles in terms of internal simulations seems to be easy. Even if “renate” and “cordate” refer to the same phenomenally accessible individuals, different subjects as well as the same subjects associate different simulations with kidneys and hearts on several occasions. In phenomenal terms alone, it is quite evident that different subjects have different information which they make use of on certain occasions in order to understand linguistic tokens. The fact that co-referential expressions cannot be substituted *salva veritate* seems to be rather trivial from the viewpoint of embodied cognition. It is a quite superficial point that a subject does not have to know that Kaka is a Brazilian football player who scored seventy Serie A goals and twenty-three Champions League goals for AC Milan in order to know that Kaka changed from AC Milan to Real Madrid. He simply does not undergo an embodied simulation in which Kaka as identical individual is simultaneously associated with the one who changed club and the one who scored so-and-so many goals. What is made clear by the hypothetical scenario of the subject simulating the information about the number of goals so that not Kaka but someone else scored them is that the referents themselves are embedded in the context of perceptual simulation. Neither a description beyond the perceptual simulation is required, nor does it make sense to extract the individuals. Since the individuals are always embedded in concrete simulations (at least in the case of offline cognition), there are two different phenomenally simulated referents in (1) and (2). The only claim that is right about the possible worlds explanation is that the subject probably does not believe the content of sentence (2). Fourthly, even if one defines possible worlds (including associated descriptions that have other referents compared to those in the actual world) in terms of scenarios which individual subjects can conceive, which seems reasonable, the fact that “Kaka” and the description in (2) have different meanings has nothing to do with the fact that they have different referents in at least one possible world, but is the result of differing perceptual simulations that were associated with the different linguistic tokens.

As previously announced, the following section aims at eliminating or reformulating classical semantic notions.

6.8 Eliminating or reformulating classical semantic notions: truth, reference, compositionality, inferential relations, and normativity

To flesh out the introductory quotation of Lakoff and Johnson, I would like to make a contribution towards rethinking “the philosophy of language from the ground up”. This includes an elimination of reference in the case of environmentally detached offline simulation, a reformulation of truth conditions most widely independent of language, and an elimination of a computational notion of compositionality and of normativity in the sense of being constitutive of mental content.

(a) What does a truth-conditional semantics purport? Basically, a theory of meaning can generate for every sentence of the object-language a T-sentence that specifies the meaning of each sentence in terms of the conditions under which it is true. For example, the German object-language sentence, “Schnee ist weiß” is true iff snow is white. The meaning of the object-language sentence is analysed in terms of a meta-language sentence.

In what way does this theory fall short? First and foremost, it does not make sense to specify the meaning of sentences without considering the context (both environmental and in the sense of prior experiences that are manifest in a state space) in which cognitive systems are embedded. Second, the fact that we understand linguistic tokens (e.g., questions, commands, metaphors) that do not have truth-values or truth-conditions as well as fact-stating sentences (cf. Lycan 2000, p. 140), should give us cause to ask for explanations of how cognitive systems understand in principle. Thirdly, the right-hand side of the T-sentence is falsely handled, as if it were an extensional description of a world that is independent of our specific human conceptual system in the way it is described. To put it pointedly: Since T-sentences unintentionally presuppose that their right side is already meaningful, they are circular.

What concept of truth condition can be formulated within the frame of embodied cognition? For a naturalistic and an anti-realistic view of the mind, it is not difficult to grant truth-conditions. Truth-conditions are specified in terms of transparent and opaque (e.g., emotions) phenomenal representations that are neurally constructed and so part of phenomenal world models. Of course, this concerns only the everyday sense of truth-conditions. Remember that we already defined truth in terms of understanding the relation between phenomenal simulations and phenomenal representations.

In what sense does it make sense to speak of understanding sentences in terms of knowing their truth conditions? In order to be able to know the meaning of a sentence one has to know

the sensorimotor simulations or the phenomenal representations (in the case of environmentally embedded usage of language, such as deixis) with which the sentence's arbitrary shapes are associated. Knowing the sensorimotor simulations with which the arbitrary shapes are associated means that your mental system almost automatically links the arbitrary shapes (to be understood as external stimuli) to corresponding sensorimotor simulations. If one assumes that social practices determine which arbitrary shapes are associated with which sensorimotor simulations, that is to say that social practices determine the truth conditions of sentences, then the associated sensorimotor simulations are true iff they fairly correspond to those that are socially stipulated.

There are at least three cases of misrepresentations⁹⁰ which are easily integrable into our sense of embodiment.⁹¹ Firstly concepts can be wrongly applied to phenomenal representations. If context-specific top-down influences make phenomenal representations appear in a way which suggests classifying them in terms concepts people would not take into account under normal conditions, then one can say that this subject conceptually misrepresents the objects of the phenomenal representations. Thus the applied simulators do not correspond to phenomenal representations. Secondly, non-phenomenally cognitively processed contents can also be cases of misrepresentations – expectations of subjects can bring forth phenomenal presentations or representations (e.g., hallucinations) that would not be triggered by the environment in the case of subjects who are not affected by hallucinations. They may be considered as cognitive, since memorized representations become merged with bottom-up information to generate a phenomenal perception. Thirdly, there are even cases of non-cognitive misrepresentations, namely, illusions. For instance, representing the Kanisza figure as consisting of two triangles and three fully filled disks is a misrepresentation, if one determines the self-generated stimulus in terms of physical structures being specified from the viewpoint of third-person phenomenality.

Summarizing, mappings of simulators onto phenomenal representations, anomalous phenomenal presentations or representations (e.g., hallucinations), and presentations and representations triggered by bottom-up stimuli (e.g., including illusions) can be cases of misrepresentations. (cf. also Siewert 2002)

(b) Now we are going to deal with the idea of reference.

As already shown, all that remains of reference is concrete reference, that is, the mapping of simulations onto phenomenal representations. It has been mentioned several times already

⁹⁰ The understanding of representation in 'misrepresentation' is theoretically unspecified.

⁹¹ The following considerations may be considered as a reply to the classical objection that naturalistic theories per se cannot account for misrepresentations.

that environmentally decoupled offline simulation does not refer to external entities at all. If you will, linguistic forms refer to sensorimotor simulations, but this happens also entirely within the cognitive systems. Well, what are the objections we have to a classical understanding of reference?

Now let us consider an argument by Putnam (1981) aimed at showing that one objectively correct account of classical reference is impossible: (P1) In a model of objectivist semantics the reference relation *R* consists of a set of ordered pairs which are made up of formal linguistic elements that are interpreted by one or more entities of the model. (P2) What does the expression 'refer' refer to? (P3) If one determined 'refer' theoretically, as in Kripke's causal theory of reference, one would again get formal sentences of a theory which would have to be endowed with meaning by *one* objective model. (P4) Yet, such an objective model cannot be possible, given that one adheres to truth-conditional semantics and to the condition that changing the meaning of the sentence's parts is sufficient for changing the meaning of the whole sentence (see the reconstruction of Putnam's argument in my subsequent remarks on compositionality).⁹² (P5) The game can be infinitely continued – one can ask for a theory of theory, of which we again need one correct interpretation, and so forth. (P6) Hence, an infinite regress threatens.

Furthermore, I think that the argument indirectly calls attention to a point that is of paramount importance for our purposes: In the context of those phenomenal entities, it is only reasonable to relate to reference if the reference relation *itself* is processed within the cognitive system. The infinite regress can only be stopped if the relata are internally processed – that is, if subjects perceptually simulate word tokens that are statistically correlated with certain content bearing sensorimotor simulations or if they map words or simulators onto phenomenal representations. Thus both the referents and the reference relations is internally processed.

(c) Now, we are concerned with the classical view of compositionality: For every complex expression *e* in *L*, the meaning of *e* in *L* is determined by the structure of *e* in *L* and the meanings of the constituents of *e* in *L*.⁹³

⁹² The point is that the truth value of sentences underdetermines the referents of their components.

⁹³ It might be that compositionality and intersubjective sharing are two dogmas which date back to Frege (Clark & Prinz 2004, p. 62). Moreover, it is not unimportant to bear in mind that classical semantics also speaks of a context principle, according to which the meanings of the sentence's parts are determined by its whole meaning. An embodied concept of semantics has the advantage that it can *explain* the context-specific nature of the meaning of linguistic expressions, be it particular or complex expressions. Clearly, this context specificity is a consequence of the context-specific nature of embodied mental content. Linguistic tokens devoid of bodily and social context simply underdetermine the underlying mental content.

In what way is this notion of compositionality problematic? Firstly, a compositional syntax and semantics in the context of classical symbolism is very problematic, as already shown. Hence, if the notion of compositionality is bound to symbolism, it has to be eliminated. I think that the empirical evidence (especially the occlusion effects) suggests the neurophenomenological implausibility of a computational notion of compositionality. The combination of context-insensitive amodal symbols cannot account for the complexity of multimodal perceptual contents.⁹⁴ In respect thereof Lakoff's idea that syntactic and semantic compositionality cannot be fully productive (see 1.2) is intelligible. Secondly, from the fact that meaningful linguistic expressions superficially exhibit compositionality one cannot conclude that compositional syntax and semantics are causally efficient structures. One of the crucial insights of connectionism is to distinguish high-level phenomena from underlying mechanisms (cf. Bechtel 2008, pp. 166-9).⁹⁵ Thirdly, if one regards concepts as action-oriented capabilities, then those context-specific situated conceptualizations are reliant on emergent features. Imagine that someone cautions you with these words: "Be careful, there is a WRONG-WAY DRIVER in a TRUCK on the motorway!" Thereupon certain features might come to your mind – for example, that cars could get caught underneath the truck, that vehicles could overturn, or that the truck could lose its shipment, etc. So if concepts are supposed to facilitate action, it is favourable that they create emergent features, most likely in by means of situated conceptualizations of dynamical perceptual simulations. Moreover, it seems uncontroversial that under certain circumstances our concepts can exhibit compositionality and systematicity, for example, if we try to understand a complex expression that we do not understand in its composition (Prinz 2002).⁹⁶ Or imagine that you read an example of systematicity ("Mary loves John" - "John loves Mary") for the first time, then your phenomenal mental episode (including the attempt to memorize this arbitrary construct) exhibits systematicity at the phenomenal level during this concrete period of time. Fourthly, Putnam (1981) gives the following argument for the inadequacy of compositional-referential and truth-functional semantics: (P1) Any adequate theory of meaning must satisfy the following condition: Changing the meaning of the sentence's parts is sufficient to change the meaning of the whole sentence. (P2) The standard model-theoretic semantics defines meaning in the following way: (i) A sentence's meaning is a function that allocates a truth

⁹⁴ This claim does not presuppose the truth of concept empiricism in terms PSS, because proponents of theories of amodal symbol systems likely assume that amodal symbols are about multimodal perceptual contents.

⁹⁵ Clearly, that is not to say that the comprehension of high-level phenomena is not constrained by knowledge of lower-level processes.

⁹⁶ Hereby we perceptually simulate the contents of the particular words in order to grasp the meaning of the whole expression.

value to that sentence in each possible world. (ii) The meaning of a term (noun or noun phrase) is a function that allocates a referent (individual or kind) to that term in each possible world. (iii) The meaning of an n-place predicate is a function that allocates a referent (a set of n-tuples of entities) to that predicate in each possible world. (P3) It is possible in principle to have two different models, A and B, for a collection of sentences, so that model A makes all sentences true and model B makes all sentences true. (P4) Hence, though the parts of the sentences of both models refer to different things, their sentences have the same truth value and consequently the same meaning. (P5) Thus, within a standard model-theoretic definition of meaning, it is possible to change the meaning of the sentence's parts without changing the meaning of the whole sentence. (C) Consequently, the standard model-theoretic definition of meaning is not an adequate theory of meaning.

The argument indirectly shows that a formal model-theoretic account of entities and relations cannot endow meaningless linguistic symbols with content, since they are themselves without meaning.

(d) Consider the following view of normativity of mental content and inferential relations: A mental state *M* has a propositional content *p* only if there is a rule, or system of rules, *R* in force for *M*. (Boghossian 2003, Millar 2004)

To what extent is this claim problematic? Or, in what way does an embodied cognitive system exhibit normativity? Firstly, our question is not whether certain inferences that can be written down are justified and why, but how such inferences are internally processed and what endows them with content. The point is not whether inferential relations can be considered as normative insofar as not every inference is allowed, but how these articulated relations are mentally represented to be meaningful. Secondly, it might well be that normative relations in terms of logical syntax are grounded in preconceptual image-schematic structures. Decisively, they are only intelligible against the backdrop of neurally embodied sensorimotor structures. Hence, as already shown, a meaningless logical structure that is only sensitive to word shapes and a non-spatial scaffold exists nowhere. Moreover, to facilitate logical processing we can also make use of symbolized logical structures by means of imagery (e.g., $A \rightarrow B$). Thirdly, I think it is none too daring to claim that grounding mental content in sensorimotor processing is sufficient, therefore that normativity cannot be constitutive of mental content, given that truth, application, and assertion conditions are the

only ones that are qualified for normativity. Fourthly, much more often than not, we do not experience ourselves as drawing inferences (Metzinger 2004b).⁹⁷

6.9 Methodological consequences of embodied cognition

As we have presented and justified a theory of offline cognition and clarified the epistemological status of simulations (see the third section of chapter two), we can now consider the implications of this theory for classical philosophical methodologies, precisely because they are often limited to 'intuition pump' (Dennett 1988) and a notion of logical possibility that is tied to conceivability.

(a) Explaining the unreliability of philosophical intuitions

I am now going to repudiate the epistemic or methodological reliability of applications of philosophical intuitions. On the one hand, philosophers use intuitions to decide whether defining conditions are individually necessary and jointly sufficient (e.g., true, justified belief), or only jointly sufficient for a definiendum (knowledge). The defining conditions are adequate if the subjects asked subsume the person/situation/individual in all logically or metaphysically possible worlds in which the person/situation/individual satisfies the defining conditions, under the definiendum. Importantly, this manner of applying intuitions aims to determine what phenomena such as knowledge, mind, or rationality *are*, not what concept people have of them. In the following, this application of intuitions (I) as epistemic (E) source (S) is abbreviated as IES. Philosophers such as Gettier (1963), Putnam (1975b), Kripke (1980, p. 83-5) and Lehrer (2000, p. 187) have made use of IES.

On the other hand, philosophers consider intuitions as evidence in need of explanation. For example, that the more or less shared intuition that mental states are multiply realizable is explained abductively, for all those who share it, in terms of their functionalist concept of the mental (Ramsey's example, 2006). This usage of intuitions (I) as evidence (E) in need of explanation (E) is abbreviated as IEE.

Clearly, the IES can only be reliable if one assumes an objectivistic view of cognition. Its reliability requires that the defining conditions (in the form of concepts) are psychologically represented – namely, contextinsensitive and intersubjectively identical. Otherwise how could it make sense or be justified to apply a layman's or philosopher's intuitions in order to determine what scientifically and philosophically relevant phenomena *are*? Such a usage of intuitions could only be reliable if the background assumptions about cognition are right. As

⁹⁷ Even less do we have the phenomenal experience that we are interpreting information which perception has supplied us with.

we have seen, an intrasubjectively and epistemically (mirroring the world) objective view of cognition is untenable, both with respect to a mind-independent world and in the context of our phenomenal world model. Hence, the IES is unreliable. By the way, it was shown that intuitions vary culturally (Jackman 2005) and are unstable or context-sensitive (Swain et al. 2008).

What about the reliability of IEE? It depends on whether one views concepts – be exceptionally understood as linguistic behaviour, hereby – as coarse-grained, macroscopically unified skills (Evans 1982) which causally rely on mental mechanisms (e.g., verbalization, perceptual simulation, linking conceptual simulation with simulation of linguistic tokens) that are not in like manner specified as it macroscopically appears, or whether one treats concepts as microfunctional⁹⁸ constituents causally explaining cognition or linguistic behaviour. I have no objections to the former. It attempts to provide at best an adequate description of a pragmatically individuated process of ascription of skills. The latter, however, is committed to a meanwhile highly questionable view that goes back to Fodor and Pylyshyn (e.g., 1988), namely, 'conceptual and propositional modularity'. One might remember at least one of the already mentioned objections: If one takes the thought of interlevel constraint seriously (that is, sensorimotor processing constitutes conceptual processing), then explanations in terms of propositions are inadequate.

To sum up: IES and the causally explanatory orientation of IEE are unreliable, not least because they make untenable assumptions about the mental. The fact that philosophical intuitions are unreliable can be explained by the context-specific nature of embodied simulations and the binding of intuitions to phenomenal simulations, as mentioned above.

(b) The other methodological consequence of embodied cognition with which we are concerned has to do with three different concepts of possibility: mental, phenomenal, and theoretical possibility.

All those worlds that can be mentally simulated are mentally possible. What especially matters is that the mechanisms the respective system makes use of in order to generate and assess representational coherence are brought into line with biological or social functionality. Possessing the property of mental possibility has nothing to do with epistemic justification.

Phenomenally possible are all those worlds we can consciously simulate. This includes states of affairs conceived of in explicit planning, cognitive operations, or hallucinations. Again, what is phenomenally possible is a consequence of the specific functional profile of the

⁹⁸ This corresponds to Clark's idea of a functional organisation that is much more fine-grained in comparison to a functional organisation being provided by symbolism (1989, 1999). I think that if one fleshes out the concept of microfunctions in terms of neural subsystems (e.g., modality-specific feature maps) that constrain phenomenal experiences, then microfunctionalism is an attractive position.

respective representational architecture. If our epistemic access to possibilities consists in conceiving scenarios, then the possibilities accessed in this way are epistemically irrelevant. They are epistemically irrelevant because the conceived content is a function of sensorimotor simulations, which is bound to affordances, for example. All that is not incompatible with situations human beings can conceive of by virtue of their bodies and brains can be mentally and phenomenally simulated.⁹⁹

Theoretical possibility is a property of worlds that can be coherently described in an external linguistic medium. What I also regard as crucial is that the notion of theoretical possibility is no longer only bound to phenomenal simulations that are conceivable without contradiction (whatever is meant by). By contrast, that which is theoretically possible/necessary is confined to certain empirical assumptions – such as, is it necessary for behaving goal-directed that humans are phenomenally conscious of the object they oriented their behaviour to? This notion of logical possibility/necessity is therefore a posteriori. Since a priori thought experiments that block out empirical evidence are phenomenally grounded, and the theoretical positions accessed in that way are frequently wrong (such as, that goal-oriented behaviour requires consciousness), theory-free a priori thought experiments are epistemically unreliable. Because such a notion of possibility explicates conceptual relations of empirically (e.g., neurophenomenological) specified concepts, it corresponds to the understanding of philosophy being introduced at the beginning of the Introduction.

There are two reasons why the difference between phenomenal and theoretical possibility is especially relevant: Firstly, it seems to make sense to think that precisely those phenomenal simulations that are internally evaluated as coherent appear intuitively plausible to us. Imagining Swamp man en route to inverted earth (Tye 1998) appears intuitively plausible to us because we can it more or less coherently simulate. As already insinuated, phenomenal possibility does not entail theoretical possibility. The fact that we are able to have a coherent phenomenal simulation of a zombie (Chalmers 1996, p. 94) is philosophically rather insignificant, because it is a function of our evolved biological endowment. That we can phenomenally simulate a seemingly possible world does not entail that this world is coherently or empirically plausible describable. Secondly, conceivability is often viewed as a property of statements. Decisively, yet, linguistically expressed propositional statements are not implied by non-propositional mental or phenomenal contents. Taking the representational or simulational dynamics unfolding in the human brain seriously, including phenomenal

⁹⁹ There are three different phenomenal simulations with three different proper functions: producing representations of the actual world that are nomologically possible; generating general models of the world that are nomologically possible and biologically relevant (in planning goal-directed actions pure cognitive processes, for example); generating quasi-linguistic representations of logically possible worlds.

processes, is at odds with assuming 'propositional modularity'. Holistically, context-sensitively and non-propositionally processed phenomenal contents (e.g., imagine a zombie) do not entail statements expressed in a linguistic medium.

To avoid misunderstandings, this is not to say that quasi-symbolic representations in the form of propositions of theoretically possible worlds cannot be phenomenally simulated. Clearly, language as an external medium can be introduced and processed at the phenomenal level – what is most important, however, is that the processed linguistic tokens are themselves modal symbols, since they are dynamically and context-specifically processed in auditory or visual form. (cf. Metzinger 2003, pp. 54-60)

I think it could be quite informative to explain and determine once what is the mental content of individual subjects undergo when they are conceiving Swamp Man or a zombie. It seems more than questionable to conclude from that the fact that people linguistically confirm that they can conceive of Swamp Man that it is theoretically possible that there are beings without consciousness which are behaviourally indistinguishable from us. It is probable that the respectively processed mental contents are far from being part of a network of theoretical assumptions which could be epistemically relevant.

7 Summary

The central difference between objectivist cognitivist semantics and embodied cognition consists in the fact that the latter is, in contrast to the former, mindful of binding meaning to context-sensitive mental systems. According to Lakoff/Johnson's experientialism, conceptual structures arise from preconceptual kinesthetic image-schematic and basic-level structures. Gallese and Lakoff introduced the notion of exploiting sensorimotor structures for higher-level cognition. Three different types of X-schemas realise three types of environmentally embedded simulation: Areas that control movements in peri-personal space; canonical neurons of the ventral premotor cortex that fire when a graspable object is represented; the firing of mirror neurons while perceiving certain movements of conspecifics.

Sensorimotor coupling and the dual visual systems thesis are readily compatible.

The body schema and the body image are functionally double dissociated phenomena: The former enables the non-phenomenal control of motor behaviour, the latter corresponds to visual phenomenal presentations and representations of the body.

The central function of the forward model is to reliably predict the consequences of behavioral output on input. It therefore makes a considerable contribution towards ensuring successful online processing.

Phenomenal contents possess the property of being available for at least one of the following three processes: guided attention, cognitive processing, and behavioral control.

Mental representations re-present a virtual external world and can become globally available.

Phenomenal representations are globally available representations of a virtual external world.

Mental simulations process more or less (e.g., predicting consequences of behavioural output on input) counterfactual contents which can become globally available. Phenomenal simulations are experienced as belonging to a subject who is transparently represented.

In consequence of processing mental presentations, blindsight patients exhibit selective motor behaviour. Phenomenal presentations are not available for concept formation.

Multilevel considerations typify mechanistic explanations in cognitive neuroscience. The notion of mechanistically mediated effects as hybrids of causal and constitutive relations and the local nature of levels of mechanisms make it possible to resist Kim's argument against higher-level causes and the problem of overdetermination.

The mechanistic framework provides an adequate concept of explanation because it does justice to the norms being implicit in cognitive neuroscience. It was shown that constitutive explanatory relevance (that is, mutual manipulability) is sufficient for causal relevance and that causal relevance (analysed in terms of the manipulationist view) accounts for the implicit norms of cognitive neuroscience.

Whether mental occurrences are multiple realisable depends on the granulation of individuation, which lower-level is chosen, and the concept of function one is trying to explain. Mental occurrences are multiple realisable at many levels. Inter-theoretic constraint does not imply unique realisation, nor is it implied by it. Hence, the argument for strong psychological autonomy and the methodological argument are inconclusive.

To the extent that the specification of neural processes depends on higher-level phenomena and the whole mechanism exhibits properties that the non-organised parts do not have, higher-level phenomena are not reducible to its non-organised realisers. That is not to say that the organised neural processing being specified in relation to macroscopically observable stimuli cannot be identical with the mechanism's whole behaviour.

Furthermore, with a view to the practice of cognitive neuroscience Bechtel and McCauley's heuristic notion of type identity is justified. This understanding of type identity is compatible with the multiple realisability of mental episodes at many levels.

Succinctly, one could say that higher-level phenomena are epistemically irreducible, ontologically reducible (regardless of whether one has a concept of ontology that depends on downward looking explanations or not), and the intelligibility of identity claims is bound to multilevel explanations.

Moreover, the conclusiveness of four classical arguments against identity claims was challenged: Neurophenomenology shows that bridging the explanatory gap is within sight; Metzinger's analysis provides a functional determination of consciousness; Jackson's Mary argument does not show that phenomenal experiences are not explanatorily reducible to neural processes; the apparent contingency of brain-mind identity may be explained by the insufficiency of our current epistemic situation; Davidson's notion of token identity is hardly intelligible against the backdrop of the practice of cognitive neuroscience.

I presented an argument against extended functionalism. The basic thought was that the intelligibility of cognitive processes is uniquely realised by neural processes – given that they are specified at an appropriate level. Thus, in respect thereof, there are no two realisers (internal and external) of the same meaningful cognitive process.

Perceptual symbols as multimodal, schematic records of the neural activation underpinning online sensorimotor processing are analogously and non-arbitrarily related to perceptual processes. PSS defines concepts in terms of simulators – that is, frames which integrate perceptual symbols across a variety of category instances into a spatial organisation and enable the potentially endless generation of specific simulations. Furthermore, productivity, the implementation of propositional structure, variable embodiment, and the processing of abstract concepts are derived features of the theory.

Because the mechanistic framework provides a justified concept of explanation and PSS was identified as a mechanistic explanation, Barsalou's theory can be considered as a justified explanation.

According to Pulvermüller's theory, language consists of networks between perisylvian areas that process word forms visually and auditorily and content-bearing sensorimotor stand-ins. The example of somatotopic activation shows that motor representations are immediate, automatic, and functionally relevant and therefore not epiphenomenal.

The fact that variables affecting perceptual processes (occlusion, size, shape, modality switching, orientation) also affect conceptual processing suggests that conception makes use of modality-specific mechanisms. Even more compelling is that lesions in specific sensorimotor areas are sufficient for knowledge deficits of categories involving contents which correspond to the lesioned areas. That the representation of categories varies both

inter- and intrasubjectively has also been corroborated. In addition to that, the situated nature of conceptualizations has been experimentally confirmed.

Almost every objection to concept empiricism has been thoroughly rejected. Among the serious ones are the possibility that phenomenal simulations are mere epiphenomena, the *prima facie* problem of abstract concepts for perceptual symbol systems, and the objection that modal theories also still need to explain context-sensitivity.

Symbolism, as the older computational model, *inter alia*, suffers from the insubstantiality of conceptual and propositional modularity. Propositional attitudes can at best be captured by internal speaking at the phenomenal level. Exclusively syntax-sensitive computing of amodal linguistic forms occurs nowhere. Although connectionist models do justice to the statistical and parallel distributed character of mental processes, their disembodiment and non-explanatory nature make them appear unsatisfactory. According to the dynamical systems theory, the mind consists of an ongoing, continuous trajectory through state space that works in real-time. Actually, neither single perceptual stimuli nor discrete mental representations occur. Online cognition has to be considered as continuous non-linear inseparable sensation-cognition-action loops. That is true of environmentally detached offline cognition only insofar as it makes use of sensorimotor structures. With regard to its input, environmentally decoupled offline cognition does not rely on visual and behavioral input.

Types of referring or directed mental occurrences are non-phenomenal motor simulations, the PMIR, and non-phenomenally or phenomenally experienced mappings of simulators onto transparent or opaque (e.g., emotions) phenomenal representations.

Environmentally decoupled offline simulations do not refer to external entities (be it mind-dependent or mind-independent ones). The internally controlled re-enactment of feature maps is all we have. The processed word forms refer to sensorimotor simulations within the cognitive systems, if you will.

The crucial problem of Putnam's natural kind externalism is the point that the idea of psychophysical identical individuals with different mental contents is inconsistent within the frame of embodied cognition.¹⁰⁰ This idea is amiss if one bears in mind that mental systems have evolved internal structures in order to coordinate their behaviour in a more complex way and not represent mind-independent entities veridically. The principal problem in Kripke's reference theory of proper names is the disregard of internal processes that are required to make thoughts about individuals intelligible.

¹⁰⁰ Asserting this inconsistency does not beg the question, if one assumes that only those theories are epistemically relevant that are compatible with causally efficient processes. Explaining the content of all types of mental episodes in terms of reference to external entities does not meet this necessary condition.

The pivotal reasons to repudiate informational semantics in terms of causal co-variance are as follows: This model cannot account for internally processed contents that are not features of our phenomenal world model. The contents of environmentally decoupled offline simulations cannot be determined by causally co-varying environmental entities. Moreover, the potentially co-varying stimuli are self-generated by the brain. It also suffers from a strong content-vehicle distinction. The assumption of incipient causes and asymmetric dependence is arbitrary and ad hoc – which is no surprise in the face of a neglect of internal processes – and only anxious to preserve epistemic realism.

Because I think that linguistic expressions are only contentful against the backdrop of phenomenal representations or simulations and that these could not differ from each other provided that Jane is psychophysically identical in both contexts, I challenge the claim that Jane's mental contents differ.

Insofar as our phenomenal presentations and representations are transparent, mental episodes are wide in phenomenal terms. Insofar as our phenomenal presentations and representations are neurally constructed in order to make some contents globally available, each processed content working against the backdrop of our phenomenal world model is specified narrowly. Content-nonconceptual content is that content which is cognitively unavailable – that is, not storable in long-term memory. State-nonconceptual processes are those processes whose occurrences do not presuppose concepts (e.g., non-phenomenal or phenomenal proprioception). State-conceptual processes are those processes which only occur, if the respective mental system has concepts available (e.g., cognitively penetrated phenomenal perceptions or environmentally detached offline simulations).

If bottom-up and top-down contents are not irreconcilable, as a rule cognitive simulations and sensory presentations and representations merge to generate phenomenal perceptual representations.

Conceptual processing in terms of PSS is indeterminate for the following reasons: Re-enacting schematically extracted symbols and the qualitative coding of high-level neurons are responsible for indeterminateness. Moreover, since phenomenal visual perceptions have a virtual character and are in this respect themselves indeterminate, conceptual processing is also indeterminate. Even the evolutionarily advantageous weakly expressed phenomenality (to avoid hallucinations) of offline simulations suggests phenomenal indeterminateness.

Non-synonymous co-referential expressions and intensional contexts are better explained in terms of context-specific associations between linguistic symbols and sensorimotor simulations than in terms of diverging referents in possible worlds.

We presented three different types of misrepresentations within the frame of our notion of embodiment. Furthermore, knowing the truth-conditions of sentences was analysed in terms of undergoing those sensorimotor simulations that are socially stipulated as being linked with the corresponding sentences (to be understood as arbitrary forms).

We presented two arguments of Putnam: The one was believed to show that a compositional-referential and truth-functionalist semantics is inadequate, because it does not fulfil the condition that changing the meaning of the sentence's parts is sufficient for changing the meaning of the whole sentence. The other was supposed to demonstrate that there is no way to speak of *one* objectively correct account of a classical notion of reference. If one proceeds on the assumption that non-phenomenal neural and phenomenal processes are interdependently specified, then arguing that the processing of inferential relations is constitutive of phenomenal mental content is more than questionable. If the mind equates to the ongoing processing of sensation-cognition-action loops, then the classical computational content-operation (to be understood as compositional logical syntax) distinction has to be abandoned.

Moreover, it might well be that the understanding of logical relations is ontogenetically grounded in pre-conceptual sensorimotor structures.

Embodied cognition has the following methodological implications: IES and the causally oriented version of IEE are unreliable and this can be explained in terms of context-sensitive sensorimotor simulations. The epistemically solely relevant concept of mental, phenomenal, and theoretical possibility is theoretical possibility.

References

- Abrahamsen A, Bechtel W (2006). Phenomena and mechanisms: Putting the symbolic, connectionist, and dynamical systems debate in broader perspective. In R Stainton, ed., *Contemporary debates in cognitive science*. Oxford: Basil Blackwell.
- Adams F, Campbell A (1999). Modality and abstract concepts. *Behavioral and Brain Sciences*, 22, 610.
- Adolphs R, Damasio H, Tranel D, Cooper G, & Damasio AR (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience*, 20, 2683-90.
- Aizawa K (draft). The Massive Multiple Realization of Psychology Properties.
- Aizawa K (2009). Neuroscience and Multiple Realization: A Reply to Bechtel and Mundale. *Synthese*, 167, 493-510.
- Aizawa K, Gillett C (2007, draft). Multiple Realization and Methodology in Neuroscience and Psychology.
- Aizawa K, Gillett C (2009). The (Multiple) Realization of Psychological and other Properties in the Sciences. *Mind & Language*, 24, 181-208.
- Akins K (1996). Of Sensory Systems and the 'Aboutness' of Mental States. Reprinted In W Bechtel, R Stufflebeam, J Mundale, & P Mandick, eds., *The Neurophilosophy Reader*. MIT Press (2001).
- Aydede M (1999). What makes perceptual symbols perceptual? *Behavioral and Brain Sciences*, 22, 610-1.
- Aziz-Zadeh L, Wilson SM, Rizzolatti G, & Iacoboni M (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol*, 16, 1818-23.
- Baars B (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Baars B (1997). *In the Theater of Consciousness. The Workspace of the Mind*. Oxford: Oxford University Press.
- Barsalou LW (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, 10, 82-93.
- Barsalou LW (1993). Structure, flexibility, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In AC Collins, SE Gathercole, &

- MA Conway, eds., *Theories of Memory* (29-101). London: Lawrence Erlbaum Associates.
- Barsalou LW (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-609.
- Barsalou LW (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 18. Reprinted In H Moss & J Hampton, eds., *Conceptual representation* (513-66). East Sussex, UK: Psychology Press.
- Barsalou LW (2005). Situated conceptualization. In H Cohen & C Lefebvre, eds., *Handbook of categorization in cognitive science* (619-50). St. Louis: Elsevier.
- Barsalou LW (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-45.
- Barsalou LW (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, 364, 1281-89.
- Barsalou LW, Wiemer-Hastings K (2005). Situating abstract concepts. In D Pecher & R Zwaan, eds., *Grounding cognition: The role of perception and action in memory, language, and thought* (129-63). New York: Cambridge University Press.
- Barsalou LW, Santos A, Simmons WK, & Wilson CD (2008). Language and simulation in conceptual processing. In M De Vega, AM Glenberg, & AC Graesser, eds., *Symbols, embodiment, and meaning* (245-83). Oxford: Oxford University Press.
- Bechtel W (2008). *Mental Mechanisms: Philosophical perspectives on cognitive neuroscience*. London: Routledge.
- Bechtel W (forthcoming). The epistemology of evidence in cognitive neuroscience. In R Skipper Jr., C Allen, RA Ankeny, CF Craver, L Darden, G Mikkelsen, & R Richardson, eds., *Philosophy and the Life Sciences: A Reader*. Cambridge, MA: MIT Press.
- Bechtel W & Richardson RC (1993). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton: Princeton University Press.
- Bechtel W, McCauley RN (1999). Heuristic Identity Theory (or Back to the Future): The Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience. In M Hahn & SC Stoness, eds., *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (67-72). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bechtel W, Mundale J (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science*, 66, 175-207.

- Bechtel W, Abrahamsen A (2005). Explanation: A Mechanistic Alternative. *Studies in History and Philosophy of the Biological and Biomedical Sciences*, 36, 421-41.
- Beer RD (1995). Computational and dynamical languages for autonomous agents. In RF Port & T van Gelder, eds., *Mind as Motion. Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press/Bradford Books.
- Beer RD (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Science*, 4(3), 91-9.
- Bermúdez JL (2007). What is at stake in the debate about nonconceptual content? *Philosophical Perspectives*, 21(1), 55-72.
- Bermúdez JL, Cahen A (2008). Nonconceptual Mental Content. In *The Stanford Encyclopaedia of Philosophy*.
- Bickle J (1998). *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press.
- Bickle J (2003). *Philosophy of Neuroscience: A Ruthlessly Reductive Approach*. Dordrecht, Holland: Kluwer Academic Publishers.
- Blackmore SJ, Brelstaff G, Nelson K & Troscianko T (1995). Is the Richness of Our Visual World an Illusion? Transsaccadic Memory for Complex Scenes. *Perception*, 24, 1075-81.
- Block N (1978). Troubles with Functionalism. In W Savage, ed., *Perception and Cognition: Minnesota Studies in the Philosophy of Science*, vol. 9, 261-325. Minneapolis: University of Minnesota Press. Reprinted in Block (1980), pp. 268-305.
- Block N (1980) (ed.). *Readings in Philosophy of Psychology*, vol. 1. Cambridge, MA: Harvard University Press.
- Block, N (1986). Advertisement for a semantics for psychology. In PA French, TE Uehling, Jr., and HK Wettstein, eds., *Studies in the Philosophy of Mind, Midwest Studies in Philosophy*, Vol. 10, 615-78. Minneapolis: University of Minnesota Press.
- Block N, Fodor J (1972). What Psychological States are Not. *Philosophical Review*, vol. 81, No. 2, 159-81.
- Boghossian P (2003). The Normativity of Content. *Philosophical Issues*, 13, 31-45.
- Boroditsky L, Ramscar M (2002). The roles of body and mind in abstract thought. *Psychological Science*, 13, 185-8.
- Borreggine KL, Kaschak MP (2006). The action-sentence compatibility effect: It's all in the timing. *Cognitive Science*, 30, 1097-112.
- Bremmer F, Schlack A, Jon Shah N, Zafiris O, Kubischik M, Hoffmann KP, Zilles K, & Fink GR (2001). Polymodal motion processing in posterior parietal and premotor cortex: A

- human fMRI study strongly implies equivalences between humans and monkeys. *Neuron*, 29, 287-96.
- Brook A, Akins K (eds.) (2005). *Cognition and the Brain: The Philosophy and Neuroscience Movement*. Cambridge University Press.
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, & Freund H-J (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, 13, 400-4.
- Burge T (1979). Individualism and the Mental. In PA French, TE Uehling, & HK Wettstein, eds. *Midwest Studies in Philosophy IV*, Minneapolis: University of Minnesota Press, 73-121.
- Burge T(1986). Individualism and Psychology. *Philosophical Review*, 95, 3-45.
- Cappa SF, Perani D, Schnur T, Tettamanti M, Fazio F (1998). The effects of semantic category and knowledge type on lexical-semantic access: A PET study. *NeuroImage*, 8, 350-9.
- Chaffin R (1997). Associations to unfamiliar words: Learning the meanings of new words. *Memory & Cognition*, 25, 203-26.
- Chalmers D (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, 2(3), 200-19.
- Chalmers D (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Chalmers D (1997). Moving Forward on the Problem of Consciousness. *Journal of Consciousness Studies*, 4(1), 3-46.
- Chao LL, Haxby JV, Martin A (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, 2, 913-9.
- Chao LL, Martin A (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage*, 12, 478-84.
- Churchland PS (2002). *Brain-Wise: Studies in Neurophilosophy*. Cambridge, MA: MIT Press.
- Churchland PS, Ramachandran VS, Sejnowski TJ (1994). A critique of pure vision. In C Koch & J Davis, eds., *Large-Scale Neuronal Theories of the Brain*. MIT Press.
- Clark A (1989). *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, MA: MIT Press.

- Clark A (1993). *Associative Engines: Connectionism, Concepts and Representational Change*. Cambridge, MA: MIT Press.
- Clark A (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
- Clark A (1999). Microfunctionalism: connectionism and the scientific explanation of mental states. English version of a paper that appears in German translation In T Metzinger, ed., *Das Leib-Seele-Problem in der Zweiten Hälfte des 20 Jahrhunderts*. Frankfurt am Main: Suhrkamp.
- Clark A & Chalmers DJ (1998). The Extended Mind. *Analysis*, 58 (1), 7-19.
- Clark A & Grush R (1999). Towards a cognitive robotics. *Adaptive Behavior* 7 (1), 5-16.
- Clark A, Prinz JJ (2004). Putting Concepts to Work: Some Thoughts for the Twenty First Century. *Mind and Language*, 19 (1), 57-69.
- Cosmelli D & Thompson E (2009). Embodiment or Envatment? Reflections on the Bodily Basis of Consciousness. In J Stewart, O Gapenne, & E Di Paolo, eds., *Enaction: Towards a New Paradigm for Cognitive Science*. Cambridge, MA: MIT Press.
- Craver CF (2009). *Explaining the brain. Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon Press.
- Craver CF, Wilson RA (2006). *Realization*. In P Thagard, ed., *Handbook of the Philosophy of Psychology and Cognitive Science*. Amsterdam: Elsevier.
- Craver CF, Bechtel W (2007). Top-down causation without top-down causes. *Biology and Philosophy*, 22, 547-63.
- Cummins RE (1975). Functional Analysis. *Journal of Philosophy*, 72 (November), 741-64.
- Cummins RE (1996). *Representations, targets, and attitudes*. Cambridge, MA: MIT Press.
- Cummins RE (2000). "How Does It Work" Versus "What Are the Laws?": Two Conceptions of Psychological Explanation. In F Keil & Robert A Wilson, eds., *Explanation and Cognition*, 117-45. MIT Press.
- Damasio AR (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25-62.
- Damasio AR (1992). The selfless consciousness. *Behavioural and Brain Sciences*, 15, 208-9.
- Damasio AR, Damasio, H (1994). Cortical systems for retrieval of concrete knowledge: The convergence zone framework. In C Koch & JL Davis, eds., *Large-scale neuronal theories of the brain* (61-74). Cambridge, MA: MIT Press.
- Darden L (2006). *Reasoning in Biological Discoveries*. New York: Cambridge University Press.

- Davies TM, Hoffman DD, Rodriguez AM (2002). Visual Worlds: Construction or Reconstruction? In A Noë, ed., *Is the visual world a grand illusion (72-87)? Journal of Consciousness Studies*, Vol. 9, No. 5-6.
- Decety J, Jeannerod M, Germain M, & Pastene J (1991). Vegetative response during imagined movement is proportional to mental effort. *Behavioral Brain Research*, 34, 35-42.
- Decety J, Grezes J (1999). Neural mechanisms subserving the perception of human actions. *Trends Cogn. Sci.*, 3, 172-8.
- De Charms RC, Zador A (2000). Neural representation and the cortical code. *Annual Review of Neuroscience*, 23, 613-47.
- De Gelder B, Vroomen J, Pourtois G, & Weiskrantz L (1999). Non-Conscious Recognition of Affect in the Absence of Striate Cortex. *Neuroreport*, 10, 3759-63.
- Delacour J (1997). Neurobiology of consciousness: An overview. *Behavioral and Brain Research*, 85, 127-41.
- Dennett DC (1988). Quining qualia. In A Marcel & E Bisiach, eds., *Consciousness in contemporary science (42-77)*. New York: Oxford.
- De Vega M, Glenberg AM, Graesser AC (eds.) (2008). *Symbols and embodiment: Debates on meaning and cognition*. Oxford: Oxford University Press.
- Dretske F (1981) *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Dretske F (1988). *Explaining behavior*. Cambridge, MA: Bradford Books/MIT Press.
- Dretske F (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
- Eliasmith C (1996). The Third Contender: A Critical Examination of the Dynamicist Theory of Cognition. *Philosophical Psychology*, 9 (4), 441-63.
- Eliasmith C (2003). Moving Beyond Metaphors: Understanding the Mind for What It Is. In A Brooks, K Akins, eds., *Cognition and the Brain. The Philosophy and Neuroscience Movement*. Cambridge University Press (2005).
- Endrass T, Mohr B, Pulvermüller F (2004). Enhanced mismatch negativity brain response after binaural word presentation. *European Journal of Neuroscience*, 19, 1653-60.
- Evans G (1982). *Varieties of Reference*. Oxford: Clarendon Press.
- Farah MJ (1989). The neural basis of mental imagery. *Trends in Neuroscience*, 12, 395-9.
- Farah MJ (2000). The neural bases of mental imagery. In MS Gazzaniga, ed., *The cognitive neurosciences (2nd ed.)*. Cambridge, MA: MIT Press.
- Field H (1977). Logic, meaning, and conceptual role. *Journal of Philosophy*, 74, 379-409.
- Fodor JA (1968). *Psychological Explanation*. New York: Random House.

- Fodor JA (1975). *The Language of Thought*. New York: Crowell.
- Fodor JA (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
- Fodor JA (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.
- Fodor JA, Pylyshyn Z (1988). Connectionism and Cognitive Architecture. *Cognition*, 28, 3-71.
- Fogassi L, Gallese V, Buccino G, Craighero L, Fadiga L, & Rizzolatti G (2001). Cortical mechanism for the visual guidance of hand grasping movements in the monkey: A reversible inactivation study. *Brain*, 124, 571-86.
- Funkhouser E (2007). Multiple Realizability. *Philosophy Compass*, 2 (2), 303-15.
- Gallagher S (1995). Body Schema and Intentionality. In J Bermúdez, N Eilan, & A Marcel, eds., *The Body and the Self* (225-44). Cambridge: MIT/Bradford Press.
- Gallagher S (2005). *How the Body Shapes the Mind*. Oxford: Oxford University Press.
- Gallagher S, Zahavi D (2008). *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. London: Routledge.
- Gibson JJ (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Gallese V (2000). The Inner Sense of Action: Agency and Motor Representations. *Journal of Consciousness Studies*, 7 (10), 23-40.
- Gallese V (2003). A neuroscientific grasp of concepts: From control to representation. *Philosophical Transactions of the Royal Society of London, B*, 358, 1231-40.
- Gallese V (2005). Embodied Simulation: From Neurons to Phenomenal Experience. *Phenomenology and the Cognitive Sciences*, 4 (1), 23-48.
- Gallese V, Lakoff G (2005). The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge. *Cognitive Neuropsychology*, 22, 455-79.
- Gettier E (1963). Is Justified True Belief Knowledge? In SM Cahn, ed., *Exploring Philosophy: An Introductory Anthology* (2009). Oxford University Press.
- Gibson JJ (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Glenberg AM, Meyer M, & Lindem K (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, 26, 69-83.
- Glenberg AM, Robertson DA (1999). Indexical understanding of instructions. *Discourse Processes*, Vol. 28(1), 1-26.
- Glenberg AM, Robertson DA (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language*, Vol. 43(3), Oct 2000, 379-401.

- Glenberg AM, Kaschak MP (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558-69.
- Glenberg AM, De Vega M, Graesser AC (2005). Framing the debate. In (the same), eds., *Symbols and embodiment: Debates on meaning and cognition*. Oxford: Oxford University Press.
- Glenberg AM, Jaworski B, Rischal M, & Levin JR (2007). What brains are for: Action, meaning, and reading comprehension. In D McNamara, ed., *Reading Comprehension Strategies: Theories, Interventions, and Technologies* (221-40). Mahwah, NJ: Lawrence Erlbaum Publishers.
- Glennan S (1996). Mechanisms and the Nature of Causation. *Erkenntnis*, 44, 49-71.
- Goldstone R, Barsalou LW (1998). Reuniting perception and conception. *Cognition*, 65, 231-62.
- Graziano MSA, Reiss LAJ, Gross CG (1999). A neuronal representation of the location of nearby sounds. *Nature*, 397, 428-30.
- Hanna R, Thompson E (2000). The mind-body-body problem. *Theoria et Historia Scientiarum*, 7 (1), 23-42.
- Hardcastle VG (1999). *The myth of pain*. Cambridge, MA: MIT Press.
- Harman G (1982). Conceptual role semantics. *Notre Dame Journal of Formal Logic*, 23, 242-56.
- Harnad S (1987). Category induction and representation. In S Harnad, ed., *Categorical perception: The groundwork of cognition*. Cambridge University Press.
- Harnad S (1990). The Symbol Grounding Problem. *Physica D* 42, 335-46.
- Harnad S (1992). Connecting Object to Symbol in Modeling Cognition. In A Clark & R Lutz, ed., *Connectionism in Context* (75-90). London: Springer.
- Harnad S (1993). Problems, Problems: The Frame Problem as a Symptom of the Symbol Grounding Problem. *Psychology* 4(34) frame-problem.11.
- Harnad S (2003). The Symbol Grounding Problem. *Encyclopedia of Cognitive Science*. Nature Publishing Group/Macmillan.
- Hauk O, Johnsrude I, Pulvermüller F (2004). Somatopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301-7.
- Hauk O, Pulvermüller F (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, 21, 191-201.
- Heck RG (2000). Nonconceptual content and the space of reasons. *Philosophical Review*, 109, 483-523.

- Hegarty M (1992). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18(5), 1084-102.
- Hoffman DD (1998). *Visual Intelligence: How we create what we see*. New York: Norton.
- Horwitz B, Braun AR (2004). Brain network interactions in auditory, visual and linguistic processing. *Brain and Language*, 89, 377-84.
- Humphreys GW, Forde, EME (2001). Hierarchies, similarity, and interactivity in object recognition. Category-specific neuropsychological deficits, *Behavioral & Brain Sciences*, 24, 453-509.
- Hurley S (1998). *Consciousness in action*. Cambridge, MA: Harvard University Press.
- Hurley S (2008). The Shared Circuits Model (SCM): How Control, Mirroring, and Simulation can Enable Imitation, Deliberation, and Mindreading. *Behavioral and Brain Sciences*, 31, 11-22.
- Husbands P, Philipides A, Smith T, & O'Shea M (2001). The shifting network. Volume signalling in real and robot nervous systems. In J Kelemen & P Sosik, eds., *Advances in Artificial Life: Proceeding of the Sixth European Conference on Artificial Life*. Berlin/Heidelberg: Springer.
- Jackman H (2005). Intuitions and Semantic Theory. *Metaphilosophy*, vol. 36, no. 3, 363-80.
- Jeannerod M (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral Brain Science*, 17, 187-245.
- Jeannerod M (1999). The 25th Bartlett Lecture. To act or not to act: perspectives on the representation of actions. *Q J Exp Psychol A.*, 52(1), 1-29.
- Johnson M (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: Chicago University Press.
- Kiefer M (2001). Perceptual and semantic sources of category-specific effects: Event-related potentials during picture and word categorization. *Memory and Cognition*, 29, 100-16.
- Kim J (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.
- Kosslyn SM, Ball TM, & Reiser BJ (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 47-60.
- Kosslyn SM (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

- Kosslyn SM (2006). *The Case for Mental Imagery*. Oxford University Press.
- Kripke, Saul (1980). *Naming Necessity*. Cambridge, MA: Harvard University Press.
- Lakoff G (1987). *Women, Fire, and Dangerous Things*. Chicago, IL: University of Chicago Press.
- Lakoff G, Johnson M (1980). *Metaphors We Live By*. University of Chicago Press.
- Lakoff G, Johnson M (1999). *Philosophy in the Flesh. The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
- Lakoff G, Núñez RE (2000). *Where mathematics comes from. How the embodied mind brings mathematics into being*. New York: Basic Books.
- Lamme VAF, Roelfsema PR (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience*, 23, 571-9.
- Landauer TK (1999). Latent Semantic Analysis (LSA), a disembodied learning machine, acquires human word meaning vicariously from language alone. *Behavioral and Brain Sciences*, 22, 624-5.
- Langacker RW (1997). Constituency, dependency, and conceptual grouping. *Cognitive Linguistics*, 8, 1-32.
- Ledoux JE (1996). *The Emotional Brain*. New York, NY: Oxford University Press.
- Leopold DA, Logothetis NK (1999). Multistable phenomena: Changing views in perception. *Trends In Cognitive Sciences*, 3, 254-64.
- Levine J (1983). Materialism and Qualia. The Explanatory Gap. *Pacific Philosophical Quarterly*, 64 (October), 354-61.
- Lewis CI (1929). *Mind and the World Order*. New York: Scribner's.
- Lowenthal F (1992). Formalism for clinical observations. In F Poyatos, ed., *Advances in nonverbal communication*. Amsterdam: John Benjamins (1992).
- Lowenthal F (1999). Can handicapped subjects use perceptual symbol systems? *Behavioral and Brain Sciences*, 22, 625-6.
- Lutz A, Thompson E (2003). Neurophenomenology: Integrating subjective experience and brain dynamics in the neuroscience of consciousness. *Journal of Consciousness Studies*, 10, 31- 52.
- Lyons W (1995). *Approaches to intentionality*. Oxford: Clarendon Press.
- Machamer PK, Darden L, Craver CF (2000). *Thinking about Mechanisms*. *Philosophy of Science*, 67, 1-25.
- Mack A, Rock I (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.
- Marchery E (2007). Concept empiricism: A methodological critique. *Cognition*, 104, 19-46.

- Martin A, Wiggs CL, Ungerleider LG, & Haxby JV (1996). Neural correlates of category-specific knowledge. *Nature*, 379, 649-52.
- Mandler JM (2004). *The Foundations of Mind: Origins of Conceptual Thought*. Oxford: University Press.
- McCulloch G (2003). *The Life of the Mind: An Essay on Phenomenological Externalism*. London: Routledge.
- McDowell J (1994). *Mind and World*. Cambridge: Harvard University Press.
- Melchner L, Pallas SL, Sur M (2000). Visual behavior mediated by retinal projections directed to the auditory pathway. *Nature*, 404, 871-6.
- Metzinger T (2003). *Being No One. The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.
- Metzinger T (2004a). Conscious volition and mental representation: Toward a More Fine-Grained Analysis. In N Sebanz & W Prinz, eds., *Disorders of Volition*. Cambridge, MA: Bradford Book/ MIT Press.
- Metzinger T (2004b). Inferences Are Just Folk Psychology. *Behavioral and Brain Sciences*, 27 (5), 670.
- Metzinger T (2009). *The Ego Tunnel. The Science of the Mind and the Myth of the Self*. NY: Basic Books.
- Metzinger T, Gallese V (2003). The emergence of a shared action ontology: Building blocks for a theory. *Consciousness and Cognition*, 12, 549-71. Special Issue on *Grounding the Self in Action*.
- Millar A (2004). *Understanding People: Normativity and Rationalizing Explanation*. Oxford: Oxford University Press.
- Milner AD, Goodale MA (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.
- Milner AD, Goodale MA (2005). *Sight Unseen: An Exploration of Conscious and Unconscious Vision*. Oxford: Oxford University Press.
- Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, & Alho K (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432-4.
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, & Winkler I (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences*, 24, 283-8.

- Neisser U (1976). *Cognition and reality: Principles and implications of cognitive psychology*. New York: Freeman.
- Niedenthal PM, Brauer M, Halberstadt JB, & Innes-Ker AH (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition and Emotion*, 15, 853-64.
- Northoff G (2004). *Philosophy of the Brain: The Brain Problem*. Amsterdam: John Benjamins.
- Oehlmann R (1999). Can metacognition be explained in terms of perceptual symbols? *Behavioral and Brain Sciences*, 22, 629-30.
- Ohlsson S (1999). Selecting is not abstracting. *Behavioral and Brain Sciences*, 22, 630-1.
- O'Keefe J (1993). Kant and the Sea-Horse. In N Eilan, R McCarthy, & W Brewer, eds., *Spatial Representation: Problems in the Philosophy and Psychology of Spatial Representation*. Oxford: Blackwell (1993). [reprinted by Oxford University Press (1999), 43-64.]
- Paillard J (ed.) (1991). *Brain and space*. Oxford: Oxford University Press.
- Pascual-Leone A, Walsh V (2001). Fast Backprojections from the Motion to the Primary Visual Area Necessary for Visual Awareness. *Science*, 20, vol. 292. no. 5516, 510-12.
- Patterson K, Hodges JR (2001). Semantic dementia. In RF Thompson, JL McClelland, eds., *International Encyclopaedia of the Social and Behavioural Sciences. Behavioural and Cognitive Neuroscience Section (3401-5)*. New York: Pergamon Press.
- Peacocke C (1992). *A Study of Concepts*. Cambridge MA: MIT Press.
- Pecher D, Zeelenberg R, Barsalou LW (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychological Science*, 14, 119-24.
- Pecher D, Zwaan R (2005). *Grounding Cognition*. Cambridge: Cambridge University Press.
- Pettigrew CM, Murdoch Bruce E, Ponton CW, Finnigan S, Alku P, Kei J, Sockalingam R, Chenery HJ (2004). Automatic auditory processing of English words as indexed by the mismatch negativity, using a multiple deviant paradigm. *Ear and Hearing*, 25, 284-301.
- Pitt D (2004). The Phenomenology of Cognition, or, What is It Like to Think That P? *Philosophy and Phenomenological Research*, 69 (1), 1-36.
- Poland JS (1994). *Physicalism: The philosophical foundations*. Oxford: Clarendon Press.
- Polger TW (2004). *Natural Minds*. Cambridge, MA: MIT Press.

- Prinz JJ (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA MIT Press.
- Prinz JJ (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Prinz JJ, Barsalou LW (2000). Steering a course for embodied representation. In E Dietrich & A Markman, eds., *Cognitive dynamics: Conceptual change in humans and machines* (51-77). Cambridge, MA: MIT Press.
- Prinz JJ, Clark A (2004). Putting Concepts to Work: Some Thoughts for the Twenty First Century. *Mind and Language*, 19 (1), 57-69.
- Pulvermüller F (2008) *Grounding language in the brain*. In M de Vega, A Graesser, & AM Glenberg, eds., *Symbols and Embodiment: Debates on Meaning and Cognition* (85-116). Oxford. Oxford University Press.
- Pulvermüller F, Preissl H (1991). A cell assembly model of language. *Network Computation in Neural Systems*, 2, 455-68.
- Pulvermüller F, Shtyrov Y, Ilmoniemi RJ (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, 17, 884-92.
- Pulvermüller F, Hauk O, Nikulin VV, Ilmoniemi RJ (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, 21, 793-7.
- Pulvermüller F, Hauk O (2006). Category-specific processing of color and form words in left fronto-temporal cortex. *Cerebral Cortex*, 16, 1193-201.
- Putnam H (1960). Minds and Machines. In S Hook, ed., *Dimensions of Mind* (148-79). New York: New York University Press. [Reprinted in Putnam (1975a), pp. 362-85.]
- Putnam H (1967). Psychological Predicates. In W Capitan & D Merrill, eds., *Art, Mind, and Religion*, pp. 37-48. Pittsburgh: University of Pittsburgh Press. Reprinted as “The Nature of Mental States”, in Putnam (1975a), 429-40.
- Putnam H (1975a) Mind, Language, and Reality. *Philosophical Papers*, vol. 2. New York: Cambridge University Press.
- Putnam H (1975b). The Meaning of Meaning. In K Gunderson, ed., *Language, Mind, and Knowledge: Minnesota Studies in the Philosophy of Science*, vol. 9 (131-93). Minneapolis: University of Minnesota Press. Reprinted in Putnam (1975a), 215-71.
- Putnam H (1981). *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- Pylyshyn ZW (1980) Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences* 3, 111-69.

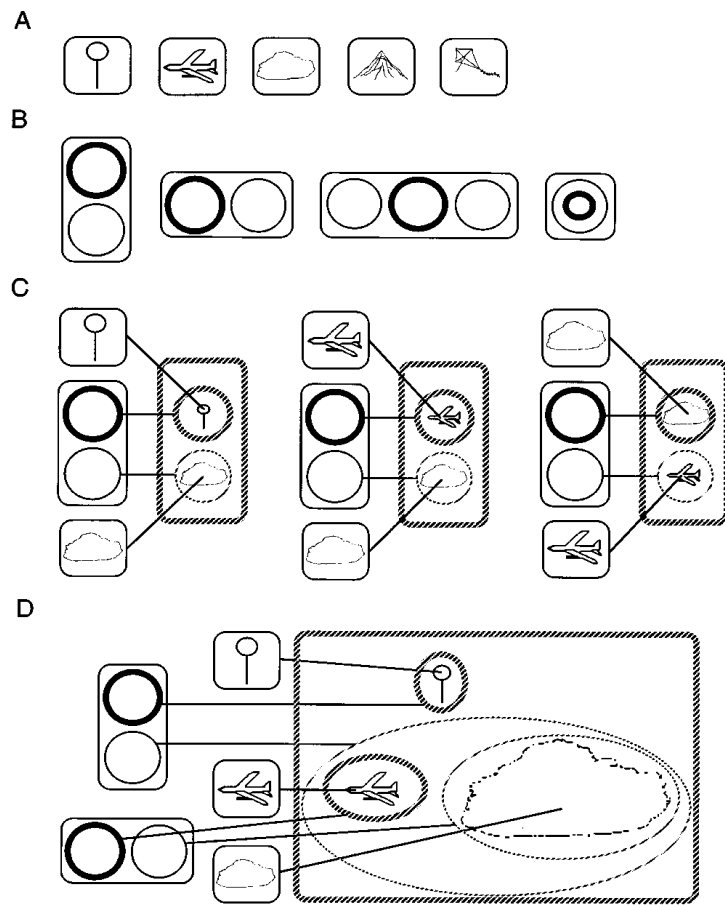
- Pylyshyn ZW (1986). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.
- Ramsey W (2006). Multiple Realizability Intuitions and the Functionalist Conception of Mind. *Metaphilosophy* 37 (1), 53-73.
- Ramsey W, Stich S, Garon J (1991). Connectionism, eliminativism, and the future of folk psychology. In W Ramsey, ed., *Philosophy and Connectionist Theory*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Rizzolatti G, Fogassi L, Gallese V (2000). Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. In MS Gazzaniga, ed., *The cognitive neurosciences* (539-52). Cambridge, MA: MIT Press.
- Rizzolatti G, Fogassi L, & Gallese V (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Neuroscience Reviews*, 2, 661-70.
- Rizzolatti G, Gallese V (2004). Do perception and action result from different brain circuits? The three visual systems hypothesis. In L van Hemmen & T Sejnowski, eds., *Problems in systems neuroscience*. Oxford: Oxford University Press.
- Rowlands M (2006). *Body Language: Representation in Action*. Cambridge, MA: Bradford Book/MIT Press.
- Rowlands M (2009). Enactivism and the Extended Mind. *Topoi* 28 (1), 53-62.
- Schmalhofer F (ed.) (2007). *Higher level language processes in the brain: Inference and comprehension processes*. Mahwah, NJ: Erlbaum.
- Schumacher R (1998). Visual perception and blindsight. The role of phenomenal qualities. *Acta Analytica*, 20, 71-82.
- Searle J (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3, 417-24.
- Sellars W (1954). Some Reflections on Language Games. *Philosophy of Science*, vol. 21, No. 3., pp. 204-28.
- Sereno SC, Rayner K, Posner, MI (1998). Establishing a time line for word recognition: evidence from eye movements and event-related potentials. *NeuroReport*, 13, 2195-200.
- Shanahan M, Baars BJ (2005). Applying Global Workspace Theory to the Frame Problem. *Cognition*, 98 (2), 157-76.
- Shapiro L (2000). Multiple Realizations. *Journal of Philosophy*, vol. 97, 635-54.
- Shapiro L (2004). *The Mind Incarnate*. Cambridge, MA: MIT Press.
- Shapiro L (2007). The Embodied Cognition Research Programme. *Philosophy Compass*, 2 (2), 338-46.

- Shaw RE (2001). Processes, Acts, and Experiences: Three Stances on the Problem of Intentionality. *Ecological Psychology*, 13, 275-314.
- Siewert C (1998). *The Significance of Consciousness*. Princeton and Oxford: Princeton University Press.
- Siewert C (2002). Is Visual Experience Rich or Poor? *Journal of Consciousness Studies*, Special Issue: Is the Visual World a Grand Illusion, vol. 9, No. 5-6, 131-40.
- Siewert C (2006). Consciousness and Intentionality. *Stanford Encyclopedia of Philosophy*. Edited by Edward Zalta.
- Simmons K, Barsalou LW (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology*, 20, 451-86. [Reprinted in A Martin & A Caramazza, eds., *The organisation of conceptual knowledge in the brain: Neuropsychological and neuroimaging perspectives* (451-86). East Sussex, UK: Psychology Press.]
- Simmons WK, Martin A, Barsalou LW (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cereb Cortex*, 15, 1602-8.
- Singer W (2000). Phenomenal awareness and consciousness from a neurobiological perspective. In T Metzinger, ed., *Neural Correlates of Consciousness. Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.
- Sober E (1992). Learning from Functionalism—Prospects for Strong Artificial Life. Reprinted in M Boden, ed., *The Philosophy of Artificial Life* (361-78). (New York: Oxford University Press, 1996). Originally, in C Langton, C Taylor, J Farmer, & S Rasmussen, eds., *Artificial Life II* (Redwood City, Calif., 1992, pp. 749-66).
- Solomon KO, Barsalou LW (1999a). Grounding property verification in perceptual simulation. (Under review).
- Solomon KO, Barsalou LW (1999b). Representing properties locally. Evidence for the perceptual grounding of concepts. (Under review).
- Solomon KO, Barsalou LW (2001). Representing properties locally. *Cognitive Psychology*, 43, 129-69.
- Spivey M (2007). *The Continuity of Mind*. New York: Oxford University Press.
- Spivey M, Tyler M, Richardson D, & Young E (2000). Eye movements during comprehension of spoken scene descriptions. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (487-92). Mahwah, NJ: Erlbaum.
- Stanfield RA, Zwaan RA (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12, 153-6.

- Stich S (1983). *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
- Strawson G (1994). *Mental Reality*. Cambridge, MA: MIT Press.
- Swain S, Alexander, Weinberg JM (2008). The Instability of Intuitions: Running Hot and Cold on Truetemp. *Philosophy and Phenomenological Research*, 76 (1), 138-55.
- Tettamanti M, Buccino G, Saccuman MC, Gallese V, Danna M, Scifo P, Fazio F, Rizzolatti G, Cappa SF, Perani D (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17, 273-81.
- Thompson E (2008). Representationalism and the Phenomenology of Mental Imagery. *Synthese*, 160 (3), 203-13.
- Thompson E, Varela FJ (2001). Radical embodiment: Neural dynamics and consciousness, *Trends in Cognitive Sciences*, 5(10), 418-25.
- Toomela A (1999). A perceptual theory of knowledge: Specifying some details. *Behavioral and Brain Sciences*, 22, 633-4.
- Tye M (1995). *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
- Vallée-Tourangeau F, Anthony SH, & Austin NG (1998). Strategies for generating multiple instances of common and ad hoc categories. *Memory*, 6, 555-92.
- Van Gelder (1998). The dynamical hypothesis in cognitive science. *Behavioural and Brain Sciences*, 21 (5), 615-28.
- Van Gulick R (1993). Who's in Charge Here? And Who's Doing All the Work? In J Heil & A Mele, eds., *Mental Causation*. Oxford: Clarendon Press, 233-56.
- Varela FJ, Thompson E, Rosch E (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.
- Wallbott HG (1991). Recognition of emotion from facial expression via imitation? Some indirect evidence for an old theory. *British Journal of Social Psychology*, 30, 207-19.
- Walter H (2001). *Neurophilosophy of Free Will: From Libertarian Illusions to a Concept of Natural Autonomy*. Cambridge, MA: MIT Press.
- Weiskopf DA (2007). Concept Empiricism and the Vehicles of Thought. *Journal of Consciousness Studies*, 14 (9-10), 156-83.
- Weiskopf DA (forthcoming). Embodied cognition and linguistic understanding. In *Studies in History and Philosophy of Science*.
- Weiskrantz L (2009). *Blindsight: A Case Study Spanning 35 Years and New Developments*. Oxford: Oxford University Press.
- Wheeler M (2005). *Reconstructing the cognitive world: The next step*. Cambridge, MA: MIT Press.

- Wheeler M (forthcoming). In Defense of Extended Functionalism. Draft of a paper to appear In R Menary, ed., *The Extended Mind*. MIT Press.
- Wheeler M & Clark A (1999). Genic Representation: Reconciling Content And Causal Complexity. *British Journal For The Philosophy Of Science*, 50 (1), 103-35.
- Willaschek M (2003). *Der mentale Zugang zur Welt: Realismus, Skeptizismus und Intentionalität*. Frankfurt am Main: Klostermann.
- Wilson CD, Simmons WK, Martin A, Barsalou LW (2007). Simulating properties of abstract concepts. Manuscript in preparation.
- Wimsatt W (1997). Aggregativity: Reductive Heuristics for Finding Emergence. *Philosophy of Science* 64 (4), 372-84.
- Wittgenstein L (1953). *Philosophische Untersuchungen*. Frankfurt am Main: Suhrkamp (2006).
- Wolpert D, Doya K, & Kawato M (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London Series B*, 358, 593-602.
- Woodward J (2003). *Making Things Happen*. New York: Oxford University Press.
- Wu L, Barsalou LW (2005). Perceptual simulation in property generation, under review.
- Zahavi D (2004). Husserl's noema and the internalism-externalism debate. *Inquiry* 47/1, 42-66.
- Zahavi D (2007). The Phenomenological Tradition. In D Moran, ed., *Routledge Companion to Twentieth-Century Philosophy*. London: Routledge, in press.
- Zola-Morgan SM, Squire LR (1993). Neuroanatomy of memory. *Annual Review of Neuroscience*, 16, 547-63.
- Zwaan RA, Stanfield RA, & Yaxley RH (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13, 168-71.
- Zwaan RA & Madden CJ (2005). Embodied sentence comprehension. In D Pecher & RA Zwaan, eds., *Grounding cognition: The role of perception and action in memory, language, and thinking* (224-45). Cambridge, UK: Cambridge University Press.

Appendix



(Barsalou 1999, p. 593)

It is important to emphasize that the pictures in figure A are not pictures or conscious images in a literal sense, but they visualize and in so far stand for configurations of neurons. These configurations of neurons become active in representing the information that is depicted in the illustrations. As important as, the particular drawings stand for simulators that can generate endless simulations of the shown instance.

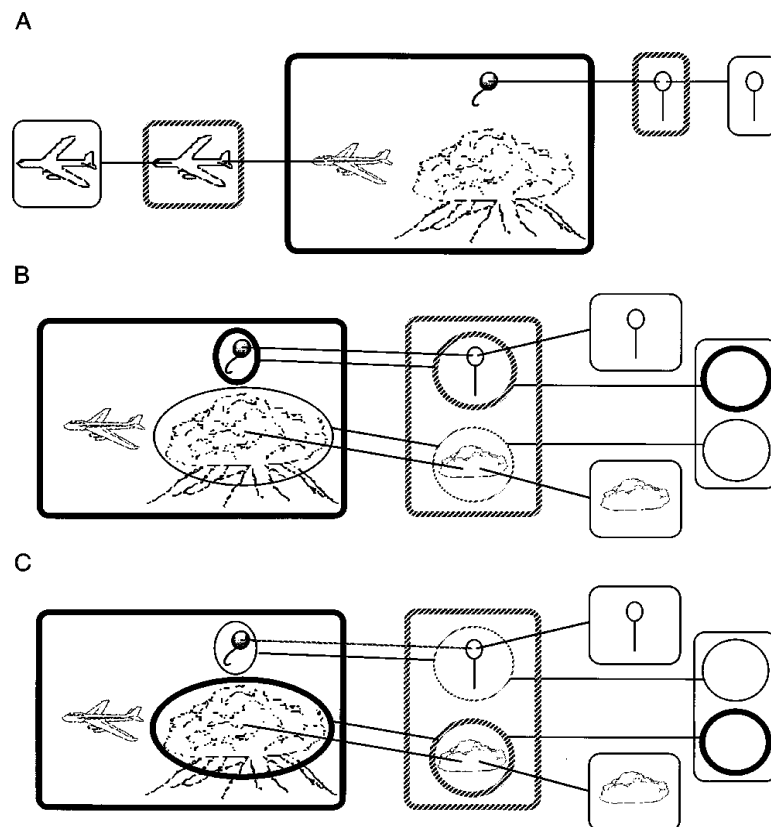
The illustrations in figure B depict simulators of spatial relations. A notion closely related to the idea of image schemata (see second section of chapter one). If you perceive a plane above a cloud, you focus on the occupied spatial regions and extract the occupying phenomenal entities. Thereupon you have a schematic representation of *above* that comprises two schematic spatial positions. Given that you filter out this information on further phenomenal perceptions, you achieve an *above* simulator that yields a variety of different *above* simulations (e.g., the two regions are vertically closer or broader apart). The thicker

boundary for a given spatial region indicates that it is that part which one's selective attention is focusing on.

In C it is depicted how simulators in figure A and B combinatorially produce complex perceptual simulations. In the rightmost example, the simulators for *above*, *cloud*, and *jet* produce a complex simulation in which a cloud is simulated above a jet.

In figure D a simulation is constructed recursively by specializing specialized schematic regions. The lower region of *above* is specialized with a *left-of* simulator's simulation, whose two regions are next specialized with simulations for *jet* and *cloud*.

Boxes with thin solid lines stand for simulators, boxes with thick dashed lines stand for simulations.



(Barsalou 1999, p. 595)

In figure A the simulator of *balloon* (type) generates a simulation of *balloon* (token) which is mapped onto or fused with the phenomenally perceived entity (token). This can be understood as the way in which PSS represents a proposition.

In figure B the simulators of *above*, *balloon*, and *cloud* produce the hierarchical simulation of a balloon above a cloud. This simulation then is fused with the phenomenal perceptual scene.

In figure C the same perceived scene like in B is alternatively interpreted. Hereby the simulator of *below* substitutes the one of *above* whereupon selective attention focuses on the cloud instead of the balloon.

Boxes with thin solid lines represent simulators, boxes with thick dashed lines stand for simulations, and boxes with thick solid lines represent perceived situations.

Zusammenfassung

Zentraler Erklärungsgegenstand dieser Magisterarbeit ist der Inhalt unserer mentalen Episoden. Die dafür in Anspruch genommenen explanatorischen Ressourcen stammen aus einem relativ jungen Paradigma namens *embodied mind*. Dieses betrachtet mentale Episoden als verkörperte bzw. neuronal verkörperte Zustände.

Das erste Kapitel beschäftigt sich mit grundlegenden Gedanken verkörperter Kognition. Darunter fällt unter anderem Lakoff/Johnsons Gedanke, dass die Inhalte höherer Formen von Kognition (wie z.B. Sprachverstehen oder Denken) anhand sensomotorischer Strukturen verstanden werden (beispielsweise verwendet man das Sprichwort „auf dem Schlauch stehen“, um eine momentane psychologische Situation einer Person intelligible zu machen). Dieser intuitiv eingängige Gedanke wird dadurch untermauert, dass bildgebende Verfahren der Hirnforschung zeigen, dass sensomotorische Gehirnareale aktiviert werden, wenn wir Sprache prozessieren bzw. über Dinge nachdenken, die nicht in der jeweils gegenwärtigen Umwelt präsent sind. Insofern spricht man auch davon, dass sensomotorische Areale von Umweltinputs entkoppelt werden können. Ihren Input erhalten sie dann aus raum-zeitlich repräsentierten Informationen, die im Langzeitgedächtnis abgespeichert sind.

Dieser naturalistische Zugang zu Kognition steht dem der kognitivistisch objektivistischen Semantik diametral entgegen. Dem letzteren zufolge haben unsere mentalen Episoden deswegen Inhalt, weil sie auf denkunabhängige Entitäten referieren. Zudem verlangt der generelle und objektive Charakter unserer quasi-sprachlichen Begriffe, dass sie unabhängig von kontextuell verkörperten Subjekten intelligible sind.

Das zweite Kapitel dient dazu, bestimmte begriffliche Unterscheidungen einzuführen, die hilfreich sind, um die Inhalte unserer mentalen Episoden umfassend beschreiben zu können. Zu Beginn dieses Kapitels wird jedoch zunächst dafür argumentiert, dass unsere phänomenalen (also bewusst erlebten bzw. erlebbaren) Wahrnehmungen (z.B. visuelle oder auditive) keinen direkten Zugang zu einer denkunabhängigen Welt verschaffen, sondern neuronal konstruierte Modelle sind, die bestenfalls einen Ausschnitt der externen Welt repräsentieren bzw. präsentieren.

Die oben erwähnten begrifflichen Unterscheidungen sehen so aus, dass Repräsentationen, Simulationen und Präsentationen von einander abgegrenzt werden. Diese werden wiederum jeweils in nicht-phänomenal und phänomenal unterteilt. Der grundlegende Unterschied zwischen Repräsentation und Simulation besteht darin, dass der Inhalt von Simulationen vom jeweiligen Erfahrungssubjekt als intern (wenn z.B. eine Person daran denkt, was sie am

nächsten Tag zu tun hat), der von Repräsentationen jedoch als extern erlebt wird (wenn man z.B. ein Auto auf der Straße sieht) (abgesehen von Halluzinationen, die als solche erlebt werden). Präsentationen zeichnen sich unter anderem dadurch aus, dass sie Informationen bereitstellen, die nicht im Langzeitgedächtnis abgespeichert werden können und daher nicht zugänglich sind für kognitives Prozessieren (wie z.B. subtile Farbnuancen). Diese drei Begriffe werden unter anderem auch dadurch charakterisiert, inwieweit sie zugänglich sind für die folgenden drei Formen mentalen Prozessierens: auf etwas seine Aufmerksamkeit richten, Kontrolle von Verhalten und kognitives Prozessieren. Diese drei Formen liefern zugleich eine funktionale Charakterisierung von Bewusstsein.

Am Ende dieses Kapitels wird dafür argumentiert, dass neuronale Vehikel und die Inhalte unserer mentalen Episoden co-konstitutiv füreinander sind.

Das dritte Kapitel zielt vor allem darauf ab, die wissenschaftliche Praxis von *cognitive neuroscience* adäquat zu beschreiben und auf dieser Grundlage klassische philosophische Themen wie das Verhältnis von Körper und Geist, Reduktionismus, mentale Verursachung oder die Frage, ob Gehirn und Geist identisch sind, zu thematisieren bzw. zu reformulieren. Einem für dieses Kapitel zentralen Gedanken zufolge integrieren neurowissenschaftliche Erklärungen verschiedene Ebenen, darunter auch die wahrnehmbaren Verhaltens. Da ich mich einer bestimmten Form von Identitätsbehauptung verschreibe, werden auch klassische Argumente gegen Identitätsbehauptungen diskutiert, allen voran Multirealisierbarkeit. Ich argumentiere unter anderem dafür, dass selbst wenn mentale Episoden multiple realisiert sind durch neuronale Prozesse, folgt daraus weder, dass Kenntnisse von Gehirnprozessen explanatorisch irrelevant sind, um mentale bzw. phänomenale Episoden zu verstehen, noch dass es keine inter-theoretischen Abhängigkeiten zwischen der Individuierung mentaler bzw. phänomenaler Episoden und neuronaler Prozesses gibt.

Das vierte Kapitel dient dazu, Barsalous Theorie perzeptueller Symbolsysteme (PSS) zu rekonstruieren. Der Erklärungsgegenstand dieser Theorie sind höhere Formen von Kognition (wie z.B. Denken oder Sprachverstehen). Ihr grundlegender Gedanke ist, dass offline Kognition sensomotorische Areale so aktiviert, dass bestimmte raum-zeitliche Inhalte nicht-phänomenal oder phänomenal prozessiert werden. Diese Theorie wendet sich vom Symbolismus ab, da amodale Symbole und ein computationales Verständnis von Propositionen (also solche, die nicht in sensomotorischen Arealen prozessiert werden) keine Rolle mehr spielen. Barsalou zeigt außerdem, dass abstrakte Begriff, wie z.B. Wahrheit, in sensomotorischen Strukturen gegründet sein können. In diesem Kapitel werden zudem empirische Evidenzen für PSS präsentiert sowie Einwände gegen PSS im Besonderen bzw.

begriffsempiristische Theorien im Allgemeinen diskutiert. Außerdem wird eine Sprachtheorie neuronaler Verkörperung dargestellt, der zufolge arbiträre sprachliche Formen nur deshalb über etwas sind bzw. Inhalt haben, weil sie mit sensomotorischen Simulationen assoziiert werden. Insofern ist das, was Sprachverstehen kausal gründet, rein intern zu fassen. Ganz im Sinne des Gedankens, dass Theorien über mentale Prozesse in dem, was neuronal passiert, gegründet sein müssen, werden im fünften Kapitel zwei computationale Modelle von Kognition (Symbolismus und Konnektionismus) kritisiert, da diese eine solche Gründung missachten. Der zentrale Einwand gegen Symbolismus ist, dass Verhalten nicht durch propositionale Einstellungen, die im Verhältnis zu quasi-linguistischen Repräsentationen stehen, erklärt werden kann. Ganz wesentlich dabei ist der Mangel an bedeutungstragenden Strukturen innerhalb der jeweiligen Subjekte. Konnektionistische Modelle stehen unter anderem vor dem Problem, dass die Modelle von *inhibition-exhibition* und *point-to-point signalling* all zu stark vereinfachen. Dieses Kapitel ist jedoch nicht rein negativ, da ein nicht-computationales Modell vorgestellt wird, namens *dynamischer Systemtheorie*. Diesem zufolge zeichnen sich mentale Prozesse durch kontinuierliche bzw. andauernde sensomotorische Aktivität aus, bei der es keine einzelnen Input-Output Computationen gibt, die nacheinander bzw. linear prozessiert werden. Dieses Modell unterscheidet sich auch dahingehend von computationalen Modellen, als es mentale Prozesse als real-zeitliche Phänomene betrachtet.

Das sechste Kapitel setzt sich zum Ziel, die philosophischen Konsequenzen von verkörperter Kognition im Generellen und Barsalous Theorie perzeptueller Symbolssysteme im Besonderen zu explizieren. Dabei wird für folgende Thesen argumentiert: (1) Die intentionale Gerichtetheit (*directedness*) mentaler Episoden besteht darin, dass Simulatoren auf phänomenale Repräsentationen projiziert werden (wenn man z.B. eine Person, die man schon einmal gesehen hat, wieder erkennt). Diese Gerichtetheit wird als konkrete Referenz verstanden. (2) In Fällen von offline Kognition gibt es keine Referenz mehr, die relevant für die Inhalte unserer mentalen Episoden ist. (3) Eine informationale Semantik (die auf der Idee kausaler Co-Varianz zwischen externer Entität und interner Episode beruht) ist vielen Problemen ausgesetzt, die mit PSS gelöst werden können. Sie kann beispielsweise nicht erklären, wie es möglich ist, dass wir über Dinge nachdenken können, die nicht in der jeweils gegenwärtigen Umwelt präsent sind. (4) Phänomenale visuelle Wahrnehmung ist kognitiv penetrierbar. (5) Das Phänomen begrifflicher Unbestimmtheit ist gut vereinbar mit PSS. (6) Zudem fassen wir den Inhalt mentaler Episoden als intern individuiert auf (*narrow content*) und wenden uns damit gegen einen sozialen oder *natural kind* Externalismus. (7) Normative

Relationen im Sinne inferentieller Beziehungen verstehen wir in Form von sensomotorischen Strukturen (*image schemata*). (8) Wenn man die traditionelle semantische Idee von Kompositionalität an ein klassisch computationales Modell von Kognition bindet, dann wird hierbei eine Elimination dieser Idee befürwortet. Wahrheit ist insofern nicht konstitutiv für den Inhalt unserer mentalen Episoden, als der Inhalt derselben in abstrakter Form bereits innerhalb der jeweiligen Subjekte vorhanden ist (das heißt, dass die Simulatoren hinreichend bedeutungstragend sind). (9) Letztlich wird die Unreliabilität philosophischer Intuitionen gezeigt und in Begriffen von *embodied cognition* erklärt.

Lebenslauf

Michael Sienhold

Bockenheimer Landstraße 135, 60325 Frankfurt am Main

- geboren am 20.08.1985 in Eisenach
 - Familienstand: ledig
-

Schulbildung

- 1992-1996 Besuch der 12. Staatlichen Grundschule in Creuzburg
- 1996-2004 Besuch des Elisabeth-Gymnasiums in Eisenach
- Abitur 2004 (Note: 2,3)

Berufserfahrung

- 01.09.04 – 31.08.05 Freiwilliges Ökologisches Jahres (FÖJ) bei der Evangelischen Jugend in Eisenach

Studium

- seit WiSe 2005/06 an der Goethe-Universität Frankfurt am Main (M.A.):
 - WiSe 2005/06 bis SoSe 2006 Philosophie (1.HF) / Allgemeine und Vergleichenden Literaturwissenschaft (2. HF)
 - seit WiSe 2006/07 Philosophie (1. HF) / Soziologie (2. HF)

Sprachkenntnisse

- Englisch
- Französisch (4 Jahre Schulunterricht)

Sonstiges

- 3. Februar 2010 Vortrag auf dem Studentenworkshop mit Michael Wheeler (zu seinem Buch „Reconstructing the Cognitive World“) (Titel: „Embodied simulation as Non-Cartesian offline cognition“)

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und dabei keine anderen als die angegebenen Hilfsmittel benutzt sowie die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, durch Angabe der Quellen kenntlich gemacht wurden.

.....

Michael Sienhold