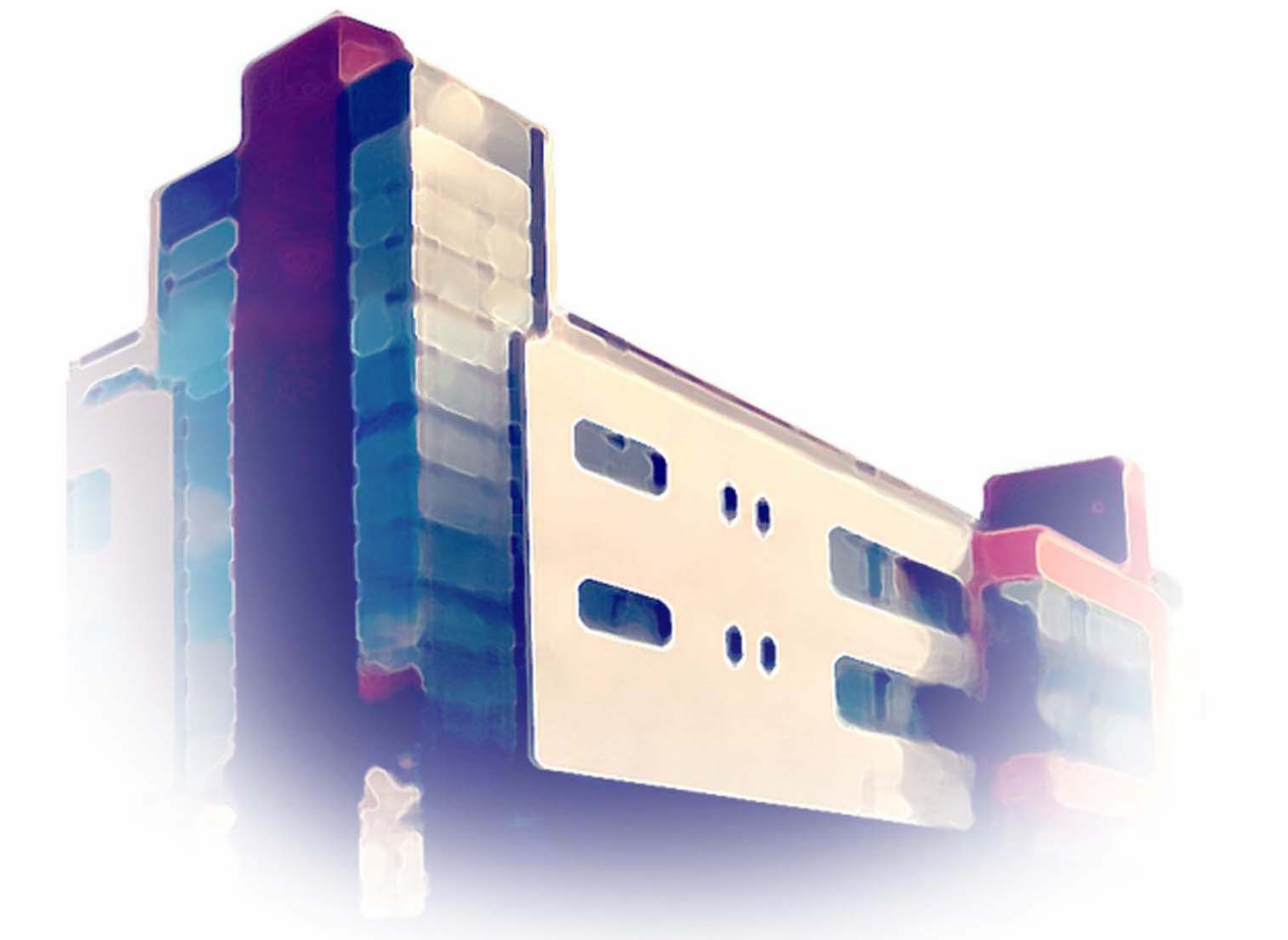
Saskia K. Nagel

Embodied, Dynamical Representationalism – Representations in Cognitive Science



PICS

Publications of the Institute of Cognitive Science

Volume 1-2005

ISSN:	1610-5389
Series title:	PICS Publications of the Institute of Cognitive Science
Volume:	1-2005
Place of publication:	Osnabrück, Germany
Date:	March 2005
Editors:	Kai-Uwe Kühnberger Peter König Petra Ludewig
Cover design:	Thorsten Hinrichs

UNIVERSITÄT OSNABRÜCK

Embodied, Dynamical Representationalism –
Representations in Cognitive Science

Saskia K. Nagel
Cognitive Science, University of Osnabrück
saskia.nagel@uos.de
Summer 2003

Contents

1. Introduction

A central problem for most explorations in the interdisciplinary science of cognition[1] is how to understand agents that behave in such sophisticated ways as they do in everyday life as well as in unusual situations requiring completely new strategies. Much of our competence depends on the ability to store information about the world, which is thought to be representations. It makes sense to think that it is the capability of representing that sets cognitive systems apart from merely complex but non-cognitive systems like oceans, ovens, organs…

In Cognitive Science we are still in quest of which paradigm to take as a foundation for exploration and explanation. To develop a framework in which the mind can be understood, a well-informed awareness about what representations are, and what problems may arise when assuming them, is very important.

In this paper I figure out what is meant when we use the notion of representation. What do we want the concept of representation for in Cognitive Science? Do we need it at all? And if so – why and in which cases? E.g. do we need it to explain each cognitive task: high-level as well as low-level?

My procedure is as follows: At first I will present how the notion is embedded in the science of cognition[2] (ch.2). This comprises what a representation is, what its inner structure looks like and how it comes to carry content. The last question is related to the symbol grounding problem which each explanatory approach is confronted with and to which it has to consider possible solutions.

To investigate these issues, I take a closer look at different modelling approaches, which constitute core methodologies in Cognitive Science: The symbolicist, the connectionist and the dynamicist as well as the embodied approach have different ideas concerning representations and the representation grounding problem.

I will describe the two classical computational approaches especially in considering their perspective on representations, which are thought to be either symbolic or distributed. I'll continue to develop their particular advantages and drawbacks (ch. 3.3.). The latter lead to a paradigm shift put forward in the field of Artificial Intelligence, often called "New AI". New AI claims a new paradigm with the core idea of body-, action- and context-dependent representations.

---

[1] I want the still vague term of "cognition" to be understood as the totality of mechanisms, which are responsible for many sophisticated behaviours such as perception, reasoning, planning, linguistic performance, and acting.

[2] I will not analyse questions such as if information-processing as the fundamental idea of Cognitive Science is a good model at all. I suppose that it is, at least it serves as a fruitful basis for understanding of what happens in processes of cognition. Nevertheless, I realize that there are groups in the field denying the information-processing paradigm by stating that one should describe and explain the brain as world-constructing vehicle.

At first, dynamicists claim that cognition is best understood as complex, dynamical, real-time interactions of a cognizer with its environment (ch.4.1.). Situatedness and embodiment add that the cognitive system is not just an encapsulated brain but instead that brain, body and environment are changing and influencing each other in a way such that system boundaries are obscure (ch. 4.2.). Although some of the proponents of these alternative approaches tend to be anti-representationalists, I'll show that they actually also need representations and in which way they do so.

At the end of my examinations I'll explain that we need representations to gain consistency with perception, which is – or at least appears to be – representational per se. I will give evidences in section 5. In this context, I will advocate a form of representationalism which does not clutch at the classical idea of a pre-existing symbol system but rather integrates the findings from connectionism, dynamicism and embodiment, and which explains our need of representation at least for some important aspects of cognition. (*Questions of priority go beyond the scope of this paper.)*

2. Role of representations in Cognitive Science

A note on history: Cognitive Science as a relatively young science has developed as a result of several important changes in the scientific field. In overcoming behaviourism later on, scientists from different research areas - today bound together in the field of Cognitive Science - constituted the so-called cognitive revolution. It was brought forward in psychology with the idea of cognitive maps in rejection of a Skinnerian behaviourism, in linguistics with Chomsky's idea about innate cognitive representational structures, and in mathematics by Turing and later McCulloch's and Pitt's frameworks for describing how input and output are interrelated. These and more perspectives taken together lead to an understanding that an adequate theory of mentality needs to posit representational states in some way between the sensory stimulus and the behavioural response.

Although it is still one of the most frequently used terms in the field, the notion of representation is one of the least clearly defined.

E.g. Cummins (1989, p.1-6) shows that there are at least four models of possible representational forms: a) as platonic ideas b) as symbols, c) as pictures and d) as neurophysiological states.

One can easily imagine that there are varieties of types of states that need to be captured by representations – some are short term, others long term, some are mediators of processes, others are modifiable, some are consciously accessible, others are not, and so on. Accordingly, as a representation, /star/[3] has to fulfil many cognitively relevant tasks: It must help the agent to

[3] From now on, I will take this notation (/xxx/) when referring to the mental representation that a cognitive agent has.

perceive stars, to remember stars, to reason about stars, to consider counterfactual situations about stars. Are the /star/s in all these cases the same, i.e. is /star/ in perception (representing states of affairs in the present), memory (representing states of affairs in the past), reasoning, and dreaming the same?[4]

Consequently, representations play a central explanatory role in the science of cognition and open a wide field for controversial discussion influenced by deep intuitions at the one hand and challenged by scientific research efforts on the other hand.

The main underlying assumption is that representational states are content-bearing and that they are supposed to carry information in some way. A system must be able to perform operations over them that lead to problem-solving and planning. Complex behaviour should be seen as being guided and controlled by inner representational states, which are patterns being activated when we perceive something.

In the next paragraph, I will examine in more detail what exactly the kind of reference or correlation relation between the representation and the represented is that should exist when we talk about representation. Let's turn to the fundamental assumptions about the concept of representation itself.

2.1. The concept of representations: Main assumptions

Our everyday experiences make us believe that some part, or feature, or constituent of an outer object is present to the subject. There is a kind of "taking-inward", which is thought to happen through the senses.[5] Obviously, not all internal entities are representations – so, by virtue of what does a representation represent, and secondly: How does a representation represent what it represents?

The traditional concept of representation is an explicit encoding of structures in the outside world leading to a correspondence or mapping relation between structures internal and external of the agent. [6]

---

[4] Here it makes sense to look at criteria for *adequacy* for a representational formalism and the world, which are pointed out by McCarthy and Hayes (1969) in Sloman (1995): *Metaphysical adequacy* is given when everything that can exist in the world can be represented. *Epistemological adequacy* is given when everything the agent needs to know can be represented. *Heuristic adequacy* is a question of relative efficiency: Purposes must be fulfilled in reasonable time by a representation. In addition, it must be considered whether the representation can be extended to cope with new developments in the environments or not. For further criteria see Sloman (1995).

[5] In contrast to opaque representations, we have to include the quite important idea of *transparency*, which has to be considered when thinking about representations. In the act of sensation our mind is directed towards the object - not towards the sensation itself. In fact, there appears to be no distinction between the experience itself, and the object that the experience is of: E.g. when I look into the sky I experience the deep blue as a property of the sky and not as a property of my perception. Although if I tried to focus on the sensation with the means of introspection, I cannot catch the phenomenon – I look just through it.

[6] For an approach to defend the notion of representation by introducing an even more fundamental concept of

To qualify some entity as representation it should satisfy several preconditions:

- Mapping: We can imagine a homomorphism[7] mapping a class of environmental structures to an internal theoretic entity. There are internal states encoding an explicit inner copy of the world in terms of law-like relations.
- Intentionality: A representation is *about* an object, state, or affair other than itself, and it is this "aboutness" that lets us distinguish it from non-representations.
- Asymmetry: The representation relation is asymmetric: If X represents Y, Y does not represent X. A painting represents a landscape but the landscape does not represent the painting. This is what distinguishes representations from mere resemblance relations, which are symmetric. If X resembles Y, Y resembles X. A person resembles another person and vice versa.
- Standing-in: As it is formulated by Haugeland (1991, p.62; italics in original): "… if the relevant features are not always present (detectable), then they can, at least in some cases, be represented; that is, something else can stand in for them, with the power to guide behavior in their stead. That which stands in for something else in this way is a *representation*; that which it stands for is its *content*; and its standing in for that content is *representing* it".

The central features are a) the *stand-in* for another (external) state, and b) the *causal power* to guide behaviour, such that it can be used by a system in order to adjust its behaviour. Thus, they coordinate behaviour possibly in absence of the represented features, in place for the direct stimuli.

To satisfy this, there must be a *representing realm* consisting of what we then call representations. The information that is being represented is in the *represented realm*. *Representing relations* determining how these two realms are correlated and *processes using the information* in the representing world allow us to assume mediating states that are internal to the cognitive system. [8]

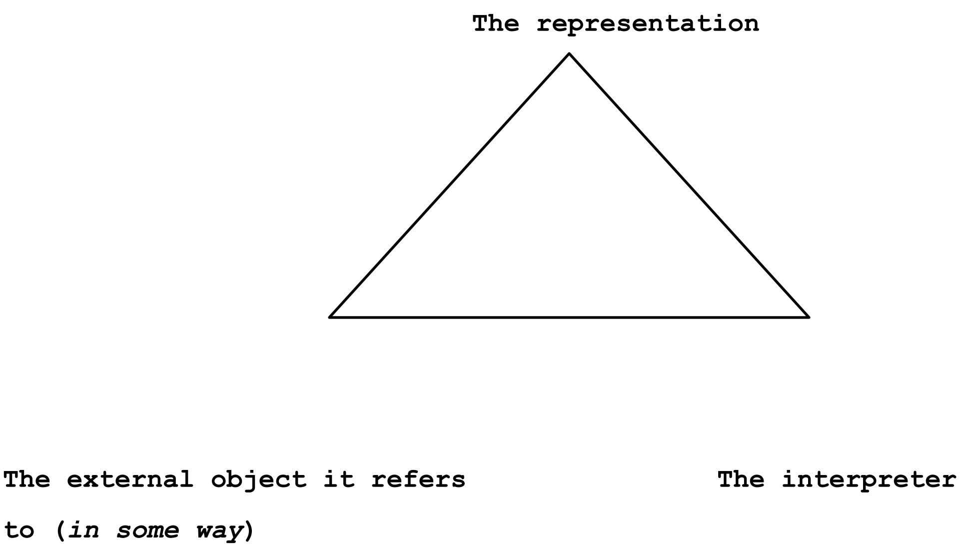
"mediating states" which are information-carrying internal states used by the system in cognitive processing, see Markman, A.B. and Dietrich, E. (1998). What should be added to mediating states to get "real" representations are five non-necessary conditions: They must be enduring, discrete, abstract, compositional and rule-governed.

[7] A short note from algebra: A homomorphism is a function which preserves the group structure: f : G → H from one group G to another group H such that f(a*b) = f(a)*f(b) for all a, b _ G. A homomorphism that is bijective (that is, both one-to-one and onto) is called an isomorphism. For a discussion concerning the identification of representations with the notion of *iso*morphism, see Haselager et al. (2003).

[8] Following Dretske (unpublished) one can distinguish the representational *vehicle*, which is the "object, event, or condition that represents" and the representational *content*, which is "the condition or situation the vehicle represents as being so". When talking about mental representations, we should be clear that the vehicle is in the head while the content is typically not. Representational content is whatever it is that constitutes a representation of something as representing exactly that. It is not the represented object in the outer world but rather the content that is represented being internal to the "system". Fred Dretske contends in his "Representational Thesis" that all mental facts are identified with representational facts. Thus, Dretske states a phenomenal *externalism* insofar as mental properties are

To be sure, reliable *correlation* between inner states and some external parameters is not enough.[9] In addition, the system should *use* the correlation, such that the system of inner states has a function in carrying the information. Neither the mere existence of complex internal states nor the existence of a correlation is enough to call a system representational, but the use and functional role to carry information in a kind of substitution function is the key here. This idea of a substitution is so wide that it includes internal processes of organisms as well as concrete things standing for other things, like a map, a morse code, or a knot in a handkerchief.

There is still another important aspect to examine – it is the *format* of a representation. According to C. S. Peirce's theory of the nature of representations (1955), a representational relation consists of the representation, the represented object and the interpretation in a cognitive system. The core idea is that a representation has a triadic structure and must be distinguished from the traditional typical dyadic relation between the representation and the represented by the additional need for an interpreter.



A representation is a representation *of* something *for* someone.

Consequently, a representation is a representation-bearer about a representational object for an

---

entirely determined by the properties they represent things as having. He explains that representational states have as main function the supply information to a cognitive system for the use in the regulation of behaviour. Dretske assumes an information-theoretical perspective, which leads to a naturalization of contents via the idea that the content of a representation can be handled with the concept of information. I'm sceptical with Dretske's assumption that information in turn is a completely objective concept that can be reduced.

[9] I.e. tide correlates with the moon position, but we would only speak of a representation if one of them stands in for the other or had evolved to carry information – but we don't do so in this case.

interpreter with the relations among the three relata. A representation is necessarily apprehended and understood by someone. Although it essentially involves an interpretation in a cognitive agent, it is not clearly defined what this interpretation looks like. Anyway, for my purpose it is sufficient to assume that the mental effects the representation has on the interpreter lead to the interpretation of it.

At this point, there is the need for a further additional feature to enable a representation to take place, which is quite intuitive and explained by Searle (1978): There is the need for a set of non-representational capacities, which is called "the background". It includes biological and cultural capacities, skills and presuppositions, i.e. information about *what* the world is *like*. To gain the right interpretation, the interpreter needs a contextual understanding – the background is a precondition for the intelligibility of a representation. We will see in section 4.2. that this idea is important in the embodied cognition approach.

I understand Peirce as well as Searle such that they explain representations on the level of the whole representational system, on a personal level. I propose that the problem with Peirce and Searle is the following: The relationship of standing-in is a relationship of transfer of representational content. In order to let such transfer occur, some entity outside must already carry the desired representational content. Hence, encodings are defined in terms of already present representations, and cannot *be* representations - cannot carry any representational content -, except via such standing-in relationships.

Thus, Peirce's and Searle's investigations do not cover *internal* representations: The observer is the source of the representational content. But what about the interpretation there? Who is the interpreter and who has the background? Questions that are hard to answer when one does not want to go back to Cartesian ideas.

2.2. How to come to content?

Simply put, the fundamental question is how an entity in the brain can represent an entity outside it. To think about that, take the following example:

A cognitive agent has the internal representation /star/. /star/ now means star, i.e. /star/ is internal to the agent and represents a star or even star-ness. /star/ may get tokened when the agent is in perceptual contact with a star and believes that there is a star. How does the external star enable the mind to have sensory perceptions of the different qualities the particular star to which it corresponds may have (as a lightening entity in the sky, a handicraft star made of wood, a star-formed cookie…)? How does it come that physical states carry content?[10]

[10] Content specifies what a representation is a representation of. It yields intentionality and representational truth-value. It is the object as it is represented to be.

There are three main approaches to answer these questions: Causal theories, conceptual role theories and two-factor-theories.

The first ones are *causal theories*: The claim is that /star/ represents star(s) because in typical or evolutionary important scenarios /star/s are *caused* by stars. As J.A. Fodor (assuming a synchronic version) and F. Dretske (assuming a diachronic version) argue - the representational relation should be seen as a causal one.[11] /star/ is about stars because it causes the token star in the head. These theories are confronted with the problem of distility (the causal chain from the stimulus to the firing cells in the brain is long, and we do not know where content arises) and the disjunction problem. The latter one is also called the problem of misrepresentations[12] and is a strong challenge for causal theories of intentional content: Representations have satisfaction-conditions, which may or may not be met. Causal theories are thought to be uncapable of capturing the possibility of mismatches between the representation and the world. If the mental token star might be caused not only by a star but also by an aeroplane, the symbol must *mean* aeroplane in the case of mismatch.[13]

There is another idea to handle the question how the brain can represent some entity from the outer world. It is *conceptual role theories*, which are also called teleological theories of mental content because of their reference to the biological role or purpose of intentional mental states. Such theories are advocated e.g. by Ned Block and show that the meaning of a representation is determined by its overall role in the conceptual scheme. They are called internalist theories of content, because they depend on factors internal to the system. /star/ means star(s) because it allows inferences about stars. The problem here is the alternate interpretation problem, i.e. that one can find an infinite number of internally consistent interpretations that are inconsistent with each other. How can there be truth of representations? In comparison to causal theories, opponents see an inability to account for truth conditions in functional theories.

The problems faced by these two kinds of theories are mutually exclusive. Thus, *two-factor-theories* combine causal and conceptual theories, supposing that this combination will help to cope with the fact that states with the same conceptual role can differ in reference when there is

[11] Dualists as well as reductionist physicalists must consider the causal status of representations as instance of the broader problem of causal efficacy of mental properties. E.g. Fodor needs a naturalized theory of mental representations when being committed to the view that intentional properties and with that representational properties are causally efficacious. It is no place here to discuss Fodor's intentional realism in length which elsewhere (Dennett, 1991) is called "industrial strength realism". Neither can I describe Dennett's weaker, "mild" intentional realism, assuming that intentional states are "somehow" real. See Dennett (1991; 1993).

[12] This problem is a variant of the argument about illusions, which describes the possibility of a mismatch between mental representation and the referent in the external world.

[13] Fodor tried to solve the problem with the introduction of asymmetric relations, which should hold between misrepresenting nomic relations and correctly representing nomic relations. There should be an asymmetry in that false evocations are dependent on true evocations but not vice versa. As far as I can see, there is no convincing approach to solve the disjunction problem yet.

another context. However, those theories are faced with the problem how the two distinct components are stuck together.

I suppose that content-fixation is realized simultaneously in different ways – thus, I want to stress the possibility to assume informational as well as functional content determination.[14] The question of content determination is densely related with the question of symbol grounding, asking how we can relate a representation to the external world intrinsically.

2.3. The connection to the external world: Grounding the representation

Traditionally, it was claimed "symbols lie at the root of intelligent action"(Newell and Simon, 1976, p.83). Thus, cognition should be understood in terms of input, output, storage and retrieval, and most importantly as manipulation of symbols. What should be manipulated is characterised as follows: Symbols are arbitrary - having intrinsic meaning or not - and they are syntactic - satisfying the criteria of compositionality and systematicity like human language.

A symbolic representation typically is completely disjoint and digital - i.e. it is always possible to determine if a token falls into a symbol class. The traditional approach seeing symbol manipulation at the core of Cognitive Science changes dramatically with the ideas of connectionism and even more with the advent of situated and embodied approaches, as we will see in the proceeding of this thesis. Each of the different approaches needs to face one crucial problem and has different ideas to solve it:

The symbol grounding problem formulated by Harnad (1990) is commencing with the observation that computation consists in the manipulation of *meaningless* symbols. Symbols acquire their meaning by projections onto them by an observer. The main question is how symbols (and in the connectionist approach the activation patterns and related subsymbolic processes) acquire their intentional content. To possess true semantic content, the symbols must somehow be grounded in a nonsymbolic base; they must be connected to the world. They cannot refer only to other symbols, but must refer to "something else" sooner or later. What might this "something else" be? As Chalmers (1992) claims that symbols should be either seen as computational token or as representation itself, I propose to use a different term for the problem. To avoid misunderstandings, it is better to talk about the *representation grounding problem*. One has to

[14] There is still another idea how to determine content proposed by neuro- and biosemantics brought in the debate mainly by R.G. Millikan (1989) who proposes that the content of a state can be manifested by looking at the evolutionary history of the mechanisms leading to this state. Of greatest importance is here the biological function the representation has for an organism. Content fixation should work like the following: A neural state means X, if the reason for the evolutionary selection is the carrying of information about X. So, if neural activation states had representational content, how would they be endowed with representational content? Do organisms without representations predate organisms without representations?

ask how a representation in a computational system can possess intrinsic content. The challenge for each approach in investigating cognitive processes is to find a vehicle for representational content that is richer than a primitive computational token. Thus, we are confronted with the question of reference and intentionality in this problem: How do grounded representations acquire intentional content? I see as a first desideratum to find out what exactly is meant when a representation is grounded. Obviously, our own mental representations *are* grounded – but how are they grounded and how can we compare their "being-grounded" to other systems?

As the representation grounding problem highlights the lack of connectedness between the representations and the represented in the external environment, we should not solely consider the *internal* aspects of the system. Later, in 4.2.2., we will see that grounding can be seen not only in subsymbolic activities but rather in direct interaction between the agent and the external world such that concepts are formed in relation to the agent's environment. This is a basis to face the problem from an embodied cognition perspective.

3. Competitive computational cognitive modeling formalisms

There are different approaches to solve the representation grounding problem. In the field of Artificial Intelligence researchers aim to gain a simulation of mental representations with computational means. Since the advent of connectionism, there is a heated dispute between the symbol system paradigm and the connectionist paradigm, the former delivering such approaches like the Turing Machine, van Neumann Machine and production systems, the latter multi-layered feed-forward networks, competitive networks and so forth. These realisations are the result of different presuppositions within the same general theory, the computational theory of the mind. This theory has two interrelated aims: a) *identifying* the structure of the system underlying our mental abilities and b) *explaining* how these abilities are gained by computations[15] within the system.

In the following paragraphs, I will describe the two different research paradigms leading to their advantages and drawbacks concerning representations as well as challenging new approaches.

3.1. Symbolist traditional treatment of representations

Since the 1950s most research in AI and Cognitive Science has been lead by the computer

[15] Computation is just the manipulation of symbol tokens on the basis of their shapes. Prototypical computational processes can be found in a Turing Machine - a machine with a tape with symbols written on it. The reading head can move the tape either forward or backward, or it can halt: It can read, write or over-write symbols. What it does is determined by what state it is currently in, what other states it can go into, and what symbol it has just read. For more detailed ideas about what computation could be and an introduction of the notion of computationalism in both paradigms, see Chalmers (1992; 1993a) and Winograd and Flores (1987).

metaphor of the mind.[16] Cognitivism (which is from here on taken as synonym for the traditional symbolist approach) tries to find the underlying primitive elements in the subject mirroring the primitive objects and relations among them in the outer world.

Accordingly, the organism carries a kind of „small-scale-model“ of external reality and its possible actions in its head. This model can be imagined as a system, which has a similar structure as the paralleled processes - the representation is described like a mirror in the inner realm of the subject. This inner realm consists of an internal model of a somehow pre-given reality, referring to external objects. In addition, there are the formally-defined and implementation-independent[17] processes which operate on the representations. Philosophers from the traditional field claim that representations are explicit. Often, mental states are described as being propositional attitudes including propositions and the agent‘s relations to them, described in linguistic forms.[18]

The classical Language of Thought Hypothesis poses that the brain must contain symbolic representations that resemble sentences of a language. Thought seems to share the properties of productivity, compositionality and systematicity with language. Thus, there are finite primitive representational states, which can be combined with recursive rules. When one tokens a representation, one thereby tokens the constituents of that representation.

The symbolist approach is densely related to - or even grounded on - the Physical Symbol System Hypothesis stating that symbol systems in a physical medium have the sufficient means for general intelligent action, which describes the full range of human intelligent behaviour. A physical symbol system is a set of entities (symbols) that are physical patterns occurring as

---

[16] This core idea leads to examples for symbolic knowledge representations as semantic nets, frames, and scripts.
Cognitive systems are approached to as digital symbol systems, which achieve intelligibility by symbolising external and internal situations and events, and by manipulating even these symbols. Even if computation and representation are not to be identified with each other, there is doubtlessly an intimate relation between the two notions in this approach. As Fodor (1975, p. 31) says: “No representations, no computations. No computations, no model.” In spite of the emergence of alternative approaches, this dominant paradigm in Cognitive Science is still defended with great fervour. Proponents state that it is "far the best theory of cognition that we've got; indeed, the only one we've got that's worth the bother of a serious discussion" (Fodor, 2000, p.1). Nevertheless, there are different opinions about this idea.

[17]As the computer-metaphor of the mind distinguishes between hard- and software, the structure on which the algorithm performs can be neglected.

[18]The approach is called the “Language of Thought Hypothesis”. It is a computational thesis about thought and thinking and is titled “the only game in town” by Fodor (1975, p.406) because it allows explaining such aspects as compositionality (illustrated by the well-known example of “John loves Mary” whose meaning is composed of its constituents’ meanings “John”+”loves”+”Mary”), productivity and the systematicity of language (if a system can think “John loves Mary” it can think “Mary loves John”). It is so called because mental representations are like sentences in a language in having a constituent structure. The hypothesis postulates a physically realized representational system with a combinatorial syntax and semantics. Operations on representations are causally sensitive to the syntactic properties. Thought is described as the tokening of representations and thinking are syntactical operations defined over representations. The Language of Thought Hypothesis is based on the Representational Theory of Mind, which in short postulates mental processes and thinking in particular as consisting of causal sequences of tokenings of mental representations. Mental representations constitute the objects of propositional attitudes.

components of expressions (symbol structures), in which they are related in a physical way. On this background, we can conceptualize processes like encoding and decoding as establishing a mapping between external world and internal representations. Computational processes operate on representation referring to the outside world. A quite important feature is that there is a *localist* correspondence between the parts of representations and the represented features: Local change in situation leads to local "modular" variation in representation following the dictum „one node - one concept".

The main drawback being at dispute - besides being too brittle - is the lack of intentionality in symbolic systems. The question of intentionality goes back to Brentano who states that it is intentionality, which distinguishes physical from mental systems where the former can never realize "aboutness".

Searle (1980) elaborates this idea in his famous and often discussed Chinese Room Thought Experiment[19], which I understand as a manifestation of the representation grounding problem. In his hotly debated 1980-paper Searle gave basics for today's ideas in so-called "new AI" when he stated that intelligence and cognition should better – if at all - be modeled by machines instead of mere programs. The classical approaches completely neglected the body as well as the environmental embedding, and concentrate research on the inner world of the subject. In the Chinese Room there is a lack of understanding due to the fact that there are no causal connections between the internal symbols and the external world they are supposed to represent. Purely symbolic AI-systems lack intentionality - they cannot relate their internal processes and representations to the external world. The question whether a robot whose structure is related to the external world has intentionality or not remains debated. If not, there must be something in addition that allows intentional states.

As another drawback, cognitivism tends to focus on higher cognitive activities such as reasoning and other language-dependent activities; low-level capacities are not captured. Furthermore, classical theory cannot explain the principle of graceful degradation of function. The degradation in symbolical systems' performance in response to damage is sudden and thus not a model for the loss of cognitive functions in humans. Neither is it able to explain holistic representation of data, nor can one realise an appreciation of context.

In summary, with respect to biological plausibility: Why should one not support the classical approach?

[19] I name it "thought experiment", not as typically traditionally done "argument", because I suppose that it is not a real argument but rather a thought experiment dealing with our intuitions that primitive computational token cannot carry intrinsic meaning.

* In symbolical systems one is confronted with the frame problem[20] arising from the questions which things change and which do not. When representing a changing world, e.g. in situation calculus, frame-axioms are needed for purported *non*changes. Otherwise the system is unable to strictly deduce that any states *persist*.
* Anatomically and physiologically the brain is a parallel system and must be seen as a highly interconnected neural structure. There is no distinct storage location. Information is addressed to content, not to location. Representations are not punctuate, but distributed.
* The hardware/software analogy fails: We have to consider the plasticity of the brain leading to a high fault tolerance.

The hypothesized parallelism between physical and symbolic events is obviously problematic. But possibly, it might still be the best explanation we yet have of how mental events can be both causal and intentional at the same time, if we are willing to assume that a robot with its relation to the outer world can in fact have intentionality.

Concerning the representation grounding problem, in the symbolist approach we only have meaningless symbols that are systematically interpretable by an external observer. Symbols have no intrinsic content as they – as far as we can say – used to have in natural systems; they are ungrounded because their meanings are mere parasites on the external interpretation.

3.2. Connectionism – another mode of computationalism

Connectionism, as a rival approach to cognitivism, provides a new framework, which is not inspired by philosophy but primarily by neuroscience. Connectionist models apparently resemble to the neurophysiological reality and therefore offer an alternative approach to overcome the aforementioned problems. The understanding of what connectionism is, varies significantly among the authors.[21] Furthermore, there is a debate about whether it is a new paradigm or merely another mode of implementation of essentially classical systems. I assume the latter perspective noticing just two different modelling approaches - as Port and van Gelder (1995, p.34) state about connectionism: “Much standard connectionist work (…) is just a variation on computationalism, substituting activation patterns for symbols”.

Connectionism does not focus primarily on problem solving and reasoning but on learning.[22] I will

---

[20] Harnad (1993) gives a helpful overview about the discussion and links the problem to the symbol grounding problem.

[21] For an introduction see Bechtel and Abrahamsen (1991), or Smolensky (1988), or Chalmers (1992; 1993b).

[22] Thus, one important source is the idea put forward by Donald Hebb (1949) by his theory of learning. When two units are activated simultaneously, the strength of the connection between them will be increased. Hence, a network can learn associations depending on correlations between activation states of units. The traditional reliance on reduction to logical primitives is challenged in this attempt by a holism, which can be already found in Wittgenstein II ideas about the role of context and purpose in the Philosophical Investigations (1953:1984). The tendency is to

now describe the characteristics of this approach in a short outline, which takes a typical net as example. Of course, there are varieties of nets with very different structures, but here this rough outline will be sufficient:

Connectionist theories model the (human) brain as a network of a number of interconnected nodes in layers: One layer receives the external input from the immediately preceding layer, another generates the output, and in between them there is a hidden layer in which the real processing is done. Nodes, which are quantitative and homogeneous, have the following properties: They have a range of different states from non-firing to firing at different intensities. They communicate these states to other nodes to which they have connections, which in turn show variation in type (they can be excitatory or inhibitory) and strength (being applicable numerically). Thus, weights are regarded to encode the information content of the whole network – they determine the response of the network to an input being supplied by providing activation to sets of units. In brevity: A single processing unit consists of a net input function defining the total signal to the unit, an activation function specifying the unit's current state, and an output function defining the signal sent by this unit to other units. Signals are sent via connections transferring numeric signals from one unit to the next and being associated with numerical strength scaling the transmitted signals.

The complexity of connectionist models can hardly be programmed in detail but must be developed by back-propagation e.g. via the generalised delta rule, or genetic algorithms adjusting the weights via a number of training trials.[23] In this context, questions of neural plausibility are discussed: Connectionists claim that anatomically and physiologically the brain is a parallel, highly interconnected system without a distinct storage location, i.e. information is encoded non-punctuate. Thus "their" form of representations, which I will introduce in the next section is thought to be neurally adequate.

In contrast to the self-perception of many connectionists, most neuroscientists doubt the biological plausibility of artificial neural nets. Following them, I just consider the oversimplification and homogenity of artificial nets, which are not biologically plausible at all.

3.2.1. Distributed – but still representations

Representations within a network can be of two different formats: *Local* representations are

---

replace the structure-sensitive processes having a computational nature by a more associativist processing with neurophysiologist basics.

[23] Connectionist networks gain their power by the process of pattern matching and generalisation. The role of categories and prototypes in the investigation of theories should not be underestimated. The question whether a particular object belongs to a concept or not can be modelled apparently more adequately by connectionist networks, because according to the non-discrete states there are continuous transitions allowing to cope better with the fuzzy nature of concepts.

(realized by) one individual unit in a certain activation state.[24] A more characteristic conception of connectionist networks is their use of *distributed* representations in contrast to local or symbolic representations favoured in the symbolists' account.[25] The standard example for the spread-outness is the concept of „grandmother“: Neurophysiologically seen, it does not seem to be represented by a single „grandmother neuron“ but rather is distributed over a whole network of neurons. What does „distributed“ mean? The item – here the grandmother – is represented by a configuration pattern over the resources, be they natural or artificial. Another item – e.g. a star – would be represented by an alternative pattern.

Within this framework the traditional idea that atomic units are meaning-bearing is rejected – the nodes themselves do not bear content, and atoms do not code for particular symbols.

In connectionism the essential idea is to use vector algebra (Smolensky, 1988; 1995) to compose complex representations with a largely implicit constituent structure. The representation of high-level conceptual entities is distributed over several nodes and one node can participate simultaneously in representing many entities. Distributed representations for complex ideas are constructed in a way that does not contain any explicit representation of their parts. Abstract representations are formed at the hidden layers and are distributed across the high-dimensional space. The relation between the representation and the to be represented is not a direct reference but an indirect generative relation.

3.2.2. Grounding distributed representations

So what about grounding in this approach? Do nets have the representation grounding problem at all? Seemingly, as a rival approach concerning the format of representation, connectionism suggests an alternative, which describes our representations grounded in distributed patterns of activity. Connectionists claim that the internal structure of a representation can systematically reflect the semantic features that it is intended to represent. Connectionist distributed

---

[24] A good example for this structure is the "interactive activation model" of reading (Rumelhart and McClelland, 1982). This network contains "word nodes" which become activated after the presentation of a particular word, and each of which represents the presence of that word. “Word nodes” are activated by "letter nodes" representing and responding to the presence of a particular letter at a particular position in the word. Finally, each “letter node” is activated by "feature nodes," representing and responding to the presence of a certain feature of the shape presented at a particular position: a horizontal/vertical bar, a curved top, etc.

[25] “Spread-out-ness” may be seen as based on the 200 year-old conflict between locationalists and holists in neuroscience. Phrenologist Franz Josef Gall claimed that different kinds of knowledge are stored explicitly in discrete brain regions, whereas Pierre Flourens argued that knowledge is distributed across at least the whole cortex. Today, there is empirical evidence for distributed encoding in the brain from monkey studies given by Schieber (1990 and 1993 discussed by Clark 1997, p.131f.) It is shown that specific cortical neurons in spatially extended activity patterns are involved in the control of several finger muscles leading to different types and directions of movements.

representations seem to be a good tool to achieve the grounding of representations in their internal structure.
But nevertheless, the relationship between an activation vector and what it is supposed to represent remains unclear. Even in connectionist models the computational primitives are simple units, which indeed themselves lack semantics - there is no epistemic content for the system itself. The systems are computational; it is just the difference that the level of computation falls below the level of representation. Concerning representational content, I assume that just adding a "learning-via-training" origin to correspondence does not help more than adding causality or lawfulness.
Representations are composed of the activities of the simple individual units but in connectionist models the atoms do not code for any symbols, they do not designate. That's why I suggest calling the activation pattern also representations – they are just in another format.
Lots of proponents of connectionism think that their systems are not computational or even that they do not need representations at all. But in fact the notion of internal representation must remain and can be fundamentally altered with the insight given by neural networks. Connectionism offers an alternative approach to understand how information might be represented in the brain: The internal distributed properties of representations carry information on what they are about.[26]
We gain a larger and much more fine-grained representational space to work with, especially because of the non-discrete nature of representation. Connectionism as a holistic approach stresses the matter of context-dependency in cognition: Any change in representative situations requires change in the whole representation – any item is stored in the context of the other items.
A big advantage of distributed encoding is that it fits overlapping storage, i.e. that one individual neuron can be part of different representations. Thus, semantically related items can be encoded by non-identical but overlapping activation patterns which helps to encode generalisations.
There are drawbacks in the connectionist attempts, too, leading to challenges for the proponents.[27]

---

[26] As an example for an impressing connectionist network the project NET-talk (Sejnowski and Rosenberg, 1986) can be considered.

[27] Fodor and Pylyshyn in their widely discussed paper (1988) claim that connectionists cannot explain compositionality and systematicity. To put it in a nutshell they pose a dilemma for connectionists: They fail to explain law-like cognitive regularities; hence they cannot provide a new paradigm. It is a challenge to adequately explain cognition without an underlying classical architecture. Connectionist models cannot represent the structure inherent to the information adequately. In fact, they can be trained to be systematic but they can also be trained not to do so. (By the way: That holds – as far as I know – for symbolist models, too.) Chalmers (1992) in reply to Fodor and Pylyshyn (1988) tries to show that connectionist systems can encode compositional structure and that they can indeed process structure-sensitive operations on distributed representations. And they can do so without

* Re-enacting the system's trajectories is quite difficult, and one cannot immediately grasp the information being comprised in it – often they seem to be even „unanalyzable“ like in the brain.
* In comparison with symbolic representations, representations in connectionism are opaque.
* *What* exactly *is* the representation in a network? Is it the pattern of activation? Is it the manner of distribution, or rather the states of nodes (such as activation levels)?

In addition, I suppose that the representation-grounding problem is not solved but rather shifted, because the question is now: How to connect symbols to lower-level sensory-motor processes to root the abstract in the concrete? To face the representation grounding problem, such a connection is absolutely necessary.

3.3. Symbolism vs. Connectionism: a short discussion

I will now come to show the basic conflict emerging from the different approaches to handle representations.

The differences in the approaches lie in the ideas about what the functional architecture of the computing system looks like, i.e. in the notions of the theory, *what kind* of computer the mind may be – an automatic formal symbol-manipulation system or a subsymbolic connectionist network. Both, classical and connectionist theories, posit representational states – both are information processing systems, relying on different mathematical foundations.

In both architectures there is a tendency to think of cognitive activities as activities occurring exclusively within the brain, which receives inputs from the outside and computes them to send back the appropriate output. Both architectures are thought to manipulate representations in some way – the question of quarrel is how this is done.

One can notice a difference in the relation between structure and process. Classicists show that the symbols in a system do not comprise themselves (intrinsically) a representational system, because they need to be manipulated by external processes that are sensitive to the symbolic structure. Connectionists claim that their systems are designed to exhibit intelligent behaviour without the need to operate with a separate processor on the structure – they do not distinguish

---

implementing classical algorithms. Symbolists assume that connectionism is a mere way of implementing symbolicists' systems. They claim that connectionist networks are only of interest to cognitive science inasmuch as they implement symbol processors. This is because it is only by physically processing physical tokens that represent the contents of mental states. Critics argued that if connectionism is mainly concerned with implementation, there is no real contribution to cognitive science, because there we need more abstract levels of description. Hence, if connectionist aspects are constrained to merely implement a symbol processor, then the interesting level of analysis for cognitive science would be symbolic, not connectionist. To cut a long story short: I assume that the brain implements a symbolic processor – it is a neural net, that's quite sure, but it is at the same time a symbolic processor at a higher level. The connectionist now should discover how symbolic processing could be build from a neural net which is more powerful than an automaton .

between data structure and the processes manipulating them.[28]

3.3.1. Benefits and costs on both sides

While the representations of symbolism are composed of discrete symbols, the representations of connectionism are not discrete regions of the model's dynamical state-space. While the rules of symbolism are as formal constraints on the symbol's relations context-free, processes in connectionism are context-sensitive constraints on the state space transitions.

There are some questions about the relationship between symbolic – high-level – representations and sensory processes: May we talk about low-level *representations* at all? Should we see symbolic representation grounded in low-level behaviour from which it obtains meaning? If so, high-level representations would be extracted out of low-level ones. There still is no satisfying idea how such a process of extraction might happen.

In symbolism, computational primitives *are* at the same time also representational primitives, they are the basic syntactic entities between state transitions and at the same time bearer of semantics. Connectionism supposes as basic bearers of meaning not simple units but patterns of activity.

Critics of the symbolist approach make us consider the fragility of symbolic systems. Confronted with partial information or situations, in which the processing system is even slightly incomplete, these systems tend to break down entirely, because symbolic representations are "all-or-none". Connectionist networks do not typically crash down when faced with slightly incomplete information or non-optimal organization in the system itself. Instead, they have the ability to exhibit what is called „graceful degradation" - they give worse, but still recognizable outputs.[29]

---

[28] In addition, Smolensky (1988, p.12) shows that subsymbolic models describe the microstructure of cognition, whereas symbolic models concentrate on its macrostructure. Consequently, I claim that the source of the classicism-connectionism debate is the different idea about what counts as cognitive: Interestingly, things we humans find effortful to do (like mathematical computation with large amounts of data), conventional computers following the classical paradigm are good at and in things we humans find effortless to do (like facial recognition) digital computers fail. Traditional models do well at high-level processes such as deductive reasoning, planning, and language use - these capacities are taken as paradigm cases. They rely on explicit, structural compositionality in which one finds component representation. Connectionist models do better at so-called low-level processes, such as pattern recognition, perception, and motor control – here they may be the appropriate models in exhibiting implicit, functional compositionality. Connectionist networks need not be pre-programmed in detail to do pattern recognition, but they learn to do so.

[29] In short: When the quality of information or internal structure degrades, the quality of the output also degrades, but the nets are still able to function quite adequately, this is *precisely because* they contain distributed representations.

Proponents of connectionism here often show up with a popular analogy: Connectionist nets seem to bear the same relation to symbolic models as accounts of quantum activity do to Newtonian explanations of "middle sized" physical phenomena. I.e. the microlevel activity can be *approximately* captured by the macrolevel descriptions, but one needs a level of description below this if one wants to understand the "real" story (Smolensky, 1988).

Prominent critics against *each* form of computationalism[30] conclude that both models – symbolist and connectionist - suffer from their abstractness and lack of embodiment. Without being embedded, connectionist models still need the intervention of the human designer helping the system in learning processes and also in generalising. In traditional connectionist networks the environment is still reduced to input and output which is provided by the designer and interpreted by the user. That means connectionists also work on cognitive phenomena as self-organised in the subject-internal part - in isolation from its many sophisticated interactions with the world.

The connection between inputs, outputs, representations and what they represent still must be done by the observer/interpreter. The representation grounding problem still waits for a solution. As in the symbolic approach, where we criticised that input and output are pre-processed and interpreted by an observer, connectionism also lacks intentionality (Searle, 1980) and real autonomy, and the human remains deeply involved in designing and constructing the system.[31]

3.3.2. Hybrid models - A way out?

I suppose that the attempts to compare the two approaches should lead towards integration. Instead of stating that either the connectionist or the symbolic level holds exclusive importance in the scientific account of cognition one should give a crucial role to each. Why should we cling to a purist version when there exists a fruitful combination? Integration is captured by hybrid models, which satisfy a compromise between both approaches in combining symbolical and connectionist compartments.[32]

One possibility for such a hybrid model is to treat connectionist models as dynamical data structures, which can be manipulated by rules. Furthermore, there is a combination of the simple building blocks of an information-processing system with natural structure-sensitive laws. The

---

[30] Like Dreyfus today and as historical precedents Heidegger in *Being and Time* (1927:2001), Merleau-Ponty in *Phenomenology of Perception* (1945:1962) and Gibson (1979), see also footnote 46.

[31] Today's models use about 100 neurons. If one can involve billions of neurons in a net, we might come nearer to intentionality and autonomy. In addition, one step in the right direction is the following: In comparison to the traditional simple feedforward networks, in recurrent networks we have - due to the relative autonomy of the system - no stable references. The hidden nodes are always already in a particular activation-state, which influences the reaction to the input and thus enables the whole system to adapt dynamically. I will soon come to an investigation of dynamics.

[32] Another alternative, which I will not present here, is the approach via control structures proposed by Aaron Sloman. He states that cognition is best conceived of as information-based control system structures for autonomous agents. Agents employ some control mechanism whose continual duty is to select the next action. Another exemplary hybrid model is the computational cognitive model CLARION, described in Sun (2000) with a top-level, which is conceptual with symbolic representations and a bottom-level, which is sub-conceptual using a back-propagation network.

implementation of such an integrated architecture typically works by taking an underlying symbolic structure and implementing it by connectionist processes. The top level consists of representations with syntactically sensitive symbolic processes performing on them; on the other side the bottom level is a connectionist model with distributed feature encoding. The encoding does not exist before the agent interacts with environment.

Learning is gained through a bottom-up process from unmediated comportment with the world to a mediated representation of it. These models save explicit representation by grounding it on experience. Smolensky (1988) presents such an integrated architecture combining features from both paradigms.[33]

Hybrid architectures try to handle the representation grounding problem such that the symbolic functions emerge as a consequence of the bottom-up grounding of categories' names in their sensory representations. Symbol manipulation is no longer only governed by the arbitrary shapes of the symbol tokens*, but by the nonarbitrary shapes of the symbol invariants in which they are grounded*. Harnad's (1990) suggestion to solve the representation grounding problem is the assumption that grounding of our representations happens internally in sensory icons. In this concept, he distinguishes the nonsymbolic iconic and categorical representations, where the formers are an "analog copy of the sensory projection" (ibid, p.8) and the latter are category-specific feature detectors, both being connected via causal connections to the objects. I think that he only shifts the problem - still the representations themselves remain meaningless. Like computationalism and connectionism, hybrid approaches also consider representations in disembodied systems. Searle (1980) argued that connectionist networks as such, without embodying them, cannot lead out of the trap of computationalism. Both paradigms as well as their combination need representations and lack a solution for the main problems concerned with that notion.

[33] He proposes an "integrated connectionist/symbolic" (ICS) architecture. Smolensky's argument is that if one is interested in the causal bases of behaviour, one must take a look at the connectionist mechanisms assumed to underlie it. The connectionist level of description doesn't give rise to a satisfactory explanation of behaviour. To gain this, one needs the symbolic level of description, which uses in fact Chomsky-style accounts of language competence. „It is often important to analyze connectionist models at a higher level; to amalgamate, so to speak, the subsymbols into symbols.“ (Smolensky 1988, p.2). Representations as distributed patterns of activity with syntactic structures are the fundamentals for a compositionally derived semantic interpretation: Patterns of network activity can realise representations with syntactic structures which are themselves not part of the representations – they should be just functionally contained in the representation. The constituent representations are "imaginary", i.e. they are patterns of activity that are constituent representations, which need not occur when the pattern of which they are constituents occur. Brian McLaughlin (1997) shows that the ICS architecture does not satisfy the principles of systematicity and productivity but is "just an implementation for LISP" and for classical Language of Thought architectures.

4. Alternative approaches – a paradigm shift

We see that just another form of computationalism is *"not sufficient"* [34], so I will investigate in the next paragraphs how different alternatives handle the question of representation in cognition. Extending the classical view needs a revolution in the eyes of many researchers in Cognitive Science. As van Gelder (1995, p. 373) puts it:

"*(...) the cognitive system is not just the encapsulated brain; rather, since the nervous system, body, and environment are all constantly changing and simultaneously influencing each other, the true cognitive system is a single unified system embracing all three*".

Thus, there are proposals how to model cognitive capacities instead. In each of the alternative approaches, the goal of representation is not an accurate mapping of the external world, but it is rather the internal structures being responsible for adequate sensori-motor transformation leading to functionally fitting behaviour. Concerning our search for an adequate idea for representation, we come to the proposal of a system-relative concept, which is strongly influenced by the environment as well as by the structure of the representational system itself, being embedded in its particular environment. [35]

The basic idea beyond is as follows: We are with our body in causal and even historical connection to the world. We are a part, a structure, of the world. Cognition therefore results from the causal relations between our structures and the external world. Somehow each of the approaches to understand the phenomenon of cognition ties representations to perceptual and effector systems with some representative mental items between them. Studies change from mere correspondence approaches, which take the mind as an isolated, more or less passive input processor to a form of pragmatism suggesting that action and interaction should be the main phenomena to focus on.

The emergence of cognitive neuroscience, and the increasing prevalence of dynamical theorizing

[34] Bickhard (1998, 2001) shows that connectionism as well as classicism is committed to a notion of representation constituted as explicit encoding. To overcome their problems, he proposes an alternative model of representation as a version of the dynamical systems approach. It is called interactivism based on the emergence of representation out of action and interaction not only of the processing of inputs. In this approach representation is implicit, unbounded, and modal – "they intrinsically involve the modality of possibility" (1998, p.6). Traditional approaches look backward in time; the interactive model in contrast is future-oriented in considering potential interactions. Interactivist representations do not "involve the notion of representation-as-correspondence-to-what-is-represented that is common to ideas of language of thought or data structures or information to be processed or connectionist "representations".“(ibid, p.21). Furthermore, he assumes that the frame problem, the problem of computational intractability, is only a problem for the standard notion of representation not for the interactive representations.

[35] For another alternative see Barsalou, L.W. (1999) who proposes perceptual symbol systems eschewing amodal representations doing justice to the fact that cognitive processing is amazingly flexible.

In addition, one should consider the "enactivist" approach from Maturana and Varela (1980) as radical critics of Cognitive Science sketching the need for a bridge between Cognitive Science and a pragmatic approach to experience via autopoietic systems.

lets dynamicists form a framework for growing amounts of work in psychophysics, perception, motor control, developmental psychology[36], cognitive psychology, situated robotics, autonomous agents research, and social psychology. The underlying hypothesis in these efforts states that cognitive agents are in the first sense dynamical systems.

In the next paragraph I take a closer look at the founding framework**.**

4.1. The concept of representation in Dynamical System Theory

Dynamical System Theory is the theoretical-mathematical approach to cognition, which is thought to have the potential to replace the classical notion of computation by a notion of cognition as an emergent property of dynamic systems, evolving in time and described in terms of topological entities (Beer, 2000).[37]

Proponents of a dynamical approach stress the physical grounding of the system in a brain, and secondly the integrated nature of the brain in a body in an environment. Cognition should be studied as a result of a complex interactive process. Dynamical System Theory is a conceptual source, which goes beyond traditional connectionism in emphasising the role of time and change. Dynamicists see cognition as an essentially temporal phenomenon, and in emphasising natural cognition as environmentally embedded, embodied and >neurally embrained<" (van Gelder, 1999, p.244), they strongly challenge the classical approaches. The core assumption is that natural cognition is not a rule-governed sequence of discrete symbolic states, but rather a dynamical phenomenon happening in time. It works according to the process of spreading activation – i.e. when a node is activated above a particular threshold it modifies the values of other nodes according to the product of its output and the weights of the connections.

A dynamical system – in short – is a set of variables changing continually and interdependently over time according to dynamical laws described by a set of non-linear difference equations. Hence, dynamical descriptions need to be deterministic, nonlinear, time-dependent, and complex by coupled differential equations. A system described in such a way represents something when

[36] See especially the work of Thelen and Smith (1994).

[37] In the last decades, there have been impressive developments in the mathematics of Dynamical System Theory, especially in the theory of nonlinear systems, complexity, and chaos. Similarly, there has been an enormous growth in available computing power, and the arrival of sophisticated programs for exploring dynamical systems. That leads to an application of dynamical theory to a wide range of natural phenomena that were previously either ignored or regarded as not being explainable. As definitions in this context, I want my reader to understand a *system* as a set of entities (variables), a *state* of the system as simply the state or value of all its variables at a time, the *behaviour* of the system as transitions between states. Furthermore, *time* in this investigation is some intrinsically ordered set, or order, serving to provide orderings over other things. The real time of systems is the set of instants at which things can actually happen, ordered by a temporal priority (before/after). Concrete events are paired with instants or periods of time, and hence stand in temporal relations with each other.

there is a correlation between the states of the dynamic system and the states of the process or behaviour to be modeled. This proposal is too vague in my opinion, but I fear that there is no better way to say it. There still is no shared idea in dynamicism about the existence and the kind of representations in dynamical systems.

Dynamical System Theory is a turn-away from the model of the mind as a discontinuous, atemporal, and purely logic-based formalism towards an emphasis of the real-time interaction of the agent with the changing world. The central problem for the traditional approach to representation is that it assumes discrete and enduring components. The new approach does not involve discrete symbols at all because it works with continuously changing variables over time and thus shows a tendency towards an abstraction with the use of collective variables.

Computationalists think of a process as starting with input and producing an appropriate output, which is generated by a system via a sequence of internal operations. Dynamicists, by contrast, think of processes as ongoing, not starting and finishing anywhere. The goal is not a mapping of an input at one time onto an output at some later time, but a constant appropriate change.

Van Gelder and Port in *Mind as Motion* (1995) claim that the Dynamical System Theory approach constitutes a revolution in sense of Kuhn's idea of paradigm change with the exclusion of symbolicism and connectionism.[38] In this new paradigm, the cognitive performance is not a discrete sequential manipulation of static representational structures, but should be seen as a structure of continuous change coevolving with the world.

Dynamicists conceive representations differently - if they assume this notion as adequate at all. Unlike digital computers[39], dynamical systems are not inherently representational. Representations need to be seen as time-locked to information in the represented external world. They are supposed to be among the entities in a dynamical system, such as parameter settings, attractors, trajectories, or bifurcation structures. Thus, I assume that Dynamical System Theory accounts are not incompatible with the other approaches, but instead should be seen as complementary. However, it is more than questionable how and if at all the concept of representations is used in this approach. Dynamical models need representations, but representations need to be reformulated as dynamical entities, which are transient and context-dependent rather than static, context-free units as in the symbolist's approach. However, the

[38] Differentiating between connectionism and Dynamical System Theory is no easy task: Connectionists assert that their models are dynamical systems, e.g. Smolensky (1988) outlines in which way a connectionist net is in fact a dynamical system. Connectionism can model cognition as the behaviour of dynamical systems and understands the networks from a dynamical perspective. On the other side dynamicists reject the connectionist commitment to computationalism and representationalism. I will not go into more detail concerning this comparison.

[39] Digital computers should be understood as systems, which carry out effective computation over representations - as systems whose behaviours are algorithmically specified finite sequences of basic operations, which constitute the manipulations of representations. They need digital variables and states as well as a time set in a discrete order.

topic of context-sensitivity would need a careful, detailed analysis, which cannot be done within this paper.

The Dynamical System approach helps us to extend our perspective: With the focus on processes evolving continuously, this approach can account for the plasticity of cognition and for individual differences (which cannot be done by classicists who focus on behaviour across individuals). In providing models of neural dynamics, Dynamical System Theory helps to understand cognition in terms of neurophysiology.

In fact, there is a group of researchers in dynamism aiming at abandoning the notion of representation completely. As van Gelder (1998, p.662) puts it: “A small but influential contingent of dynamicists have found the notion of representation to be dispensable or even a hindrance for their particular purposes.” Indeed, Dynamical System Theory seemingly has an affinity for anti-representationalism (like it is proposed by van Gelder (1995), and Thelen and Smith (1994)) in seeing cognitive agents as dynamical systems primarily tightly coupled with the environment.

Van Gelder et al. (1995) assume cognition as skilful coping with the external world without representing the world, i.e. without constructing a world-model – put differently, cognition should *transcend* representation. Many have argued that mental phenomena as representations are not relevant for the modelling of embodied autonomous agents – getting the dynamics right should be the only relevant criterion.

4.1.1. Intermediate conclusion and a first proposal

Although there are obvious phenomena being captured by the view of dynamicism, we can imagine many cognitive capacities not involving any strong coupling with the environment at all, like reasoning, imagery, and planning - but which instead can be seen as mainly internal processes. Thus, an agent can act directly with the environment *and* also with internal representations of the environment.

I now want to propose that exactly this capacity is crucial in exploring representational states: We can switch between a coupling to the *real* external environment and an internal environmental *model*. This capacity makes us such good cognizers as we are. Imagine that you as an agent are able to decouple from the external environment and switch to a coupling to the internal model. That is a really amazing capacity!

Additionally, looking for this capacity helps us to distinguish between cognitive and non-cognitive systems: The former can *selectively* work within the system or with the environment, the latter cannot.

An example of a distinction between complex behaviours, which are genuinely cognitive, from those, which are merely complex and adaptive is elaborated by Grush (1997). His basic idea is to distinguish **re**presentations from mere presentations. Cognitive behaviour depends on an emulation or simulation of some external system, that can be used off-line (without the external system being (perceivably) present), in order to provide information about what the external system would do under diverse conditions. Thus, the emulator represents the external system, and we can engage, for instance, in counterfactual reasoning about the external system. Only in the case of a *representation* there is an inner state that stands-in for some other state, and that is used by a system in coordinating its behaviour towards the object or event represented. In the case *of a presentation there is no off-line use of certain information.* Thus, what the agent acquires through experience is not *represented* in the mind, but rather is *presented* to him as a discriminated situation, which then solicits a more and more refined response. Presentations play a role in the production of behaviour, but this role is embedded in the direct and continuous coupling between the presentation and the elements presented by it.

An important question concerning this distinction is whether the stimulus must be external or not. If it might be internal, the distinction is problematic because we cannot distinguish between off- and online-usage appropriately.

This discovery shows us that we need both: The notion of representation as well as the notion of a reality we are more or less densely coupled with. Hence, we need an agent in an environment the agent can - simultaneously - represent internally. *Representations are sensible entities used to stand in for something else and having a content exactly in standing in for some real target in the external world* - just as we defined it in the beginning.

Consequently, I assume that cognition is both: a dynamical as well as a representational process.

We have seen that Dynamical System Theory provides additional alternatives to the previously shown models of representational format. The contribution this theory makes to Cognitive Science is the focus on changing processes in a system serving to carry information and representing it.

As far as I can see, it is impossible for dynamicists to give a full account of human cognition, without an account to the representational aspects of complex interactions of thinking.

A dynamical approach becomes (even) more powerful when it is coupled with the situated and embodied cognition approach.

4.2. Situatedness and embodiment as even one step further

We have seen that there is the need to describe the behaviour and the cognitive processes of an agent as outcome of a close coupling[40] between agent and environment. This is also based on evidence from other sciences within the field of cognitive science. Concerning the phenomenon of vision, which is investigated in neurophysiology, psychophysics and psychology, the necessity of a coupling is considered to be crucial. Traditionally, approaches are based on the idea that in the process of vision the brain produces an internal representation of the world. The use of the internal representations, which are oftentimes assumed to be retinotopically organized cortical maps, gives rise to the experience of seeing. However intuitive that might look like, the problem with this idea is that it cannot explain how the internal representation can produce the quality of visual consciousness. Thus, there are alternative proposals assuming that seeing is a way of acting. This should be understood as a particular way of exploring the environment – vision as exploratory activity: Thus, the mere activity in internal representations does not generate the experience of seeing. The outside world serves as its own, external representation; the world is its own model, as Clark puts it. This approach provides a natural and principled way not only to account for visual consciousness, but also for differences in the perceived quality of sensory experience in the different sensory modalities**.** O'Regan and Noe (2001) present several lines of empirical evidence supporting this theory, e.g. from experiments in sensorimotor adaptation, the phenomenon of visual stability despite eye movements, change blindness, and from experiences with sensory substitution.[41]

It is pointed out that the organism's components form together one whole - a subject - which is itself functionally embedded in its environment to a systematic whole. Concerning our idea of representations we consequently need a *physical* grounding: Building an intelligent system needs a grounding of representations in the physical world - and for this we need to connect it to the world via sensors and actuators, i.e. to lower-level sensory-motor processes. There should not be a traditional internal model of the external world like a mirror, but rather an emergence of representations via action selection. However, even this new approach presupposes some kind of features. We are confronted with a pragmatic turn – even more, Clancey (1991, p.109) calls it a "Copernican shift": "Representation are not at the centre of the mind but rather emerge from the interaction of mental processes with the environment". Accordingly, Brooks as a strong proponent of this new approach characterises embodiment and situatedness as two cornerstones

---

[40] Two systems are *structurally coupled* when they repeatedly perturb one another's structure over a period of time.

[41] Whether their proposal based on sensory-motor contingencies entails an anti-representationalist conclusion or not needs to be investitgated further at another place.

of the new approach to Artificial Intelligence[42]. For him, a system needs to be embodied to experience the world and it needs to be situated to acquire knowledge through interactions with the world. Embodied Cognitive Science came around in the late 1980s, having its roots in artificial intelligence and behaviour-based robotics, but also in psychology (Gibson, 1979), linguistics (Lakoff, 1987), and neuroscience (Edelman, 1987).

With the advent of this approach, there was a shift in the focus of attention: The generation of behaviour facilitating the organism's survival became more important than an accurate mapping between outer world and inner symbol system. The primary task of a cognitive system is the contribution to interaction. A motto of embodiment is: "At first care for the survival – then worry about the rest." "The rest" is any kind of problem solving on the internal information-processing system. There is a priority of the bodily interaction to pure thinking. (Notice that the latter may still crucially determine the former task.)

Now: How can an intelligent agent who is able to view its environment from its own perspective relate mental objects to behaviour? This problem is handled by the idea of *situatedness.*

The term "situated" is usually intended to suggest that the organism/agent does not deal with mere abstract descriptions but in the „here and now" of the actual world which directly influences its behaviour. Reasoning processes are not only conditional on the environment, but are in fact created by an interactive process among different systems. Situated reasoning is reasoning by embodied beings acting in physical environments. Intelligent behaviour cannot be described only by pre-existing internal structures, but need to consider interacting biological, physical and social processes. The idea of memory as stored representations, which correspond to objective properties of reality as well as seeing concepts as pre-defined feature lists must be rejected.

---

[42] Work involving mobile robots, especially at the MIT AI Laboratory plays a crucial role in establishing these ideas. Braitenberg's vehicles which are simple vehicles exhibiting animal-like behaviour (e.g. light-avoiding) take the external world itself not an internal replica as the spatial medium – and Brooks elevates this concept to a new principle of robotics: He and his colleagues constructed reacting robots behaving in real-time and real-world conditions by building robot bodies on the base of an behaviour-oriented architecture, the so-called "subsumption architecture" (Brooks, 1991a, 1991b). It has layered augmented finite state machines, allowing a close coupling among sensors, actuators and hardware. This idea is crucial because it is not an environment- and body- independent system but an implementation of an perception-action coupling for low-level behaviour which is extended by the addition of further layers just mediating between perception and action. The lower layer's activity can be subsumed (inhibited) if necessary. The point is that we do not need a central control, or a central planner or an inner model of the world. The world itself is used as a medium for communication between the layers. Brooks states that intelligent behaviour could be generated without having explicit manipulable internal representations, without central control. "We never use tokens which have any semantics that can be attached to them" (Brooks, 1991b, p.6). Nevertheless, he does not reject the ideas of representations at all. Rather he assumes representations as partial models: "Individual layers extract only those aspects of the world which they find relevant-projections of a representation into a simple subspace" (ibid).

Memorising and comprehending is not storing a representation but the coordination of perception and action according to new necessary adaptations. Thus, instead representations are described as being created and interpreted interactively as a result of perceiving and behaving. All processes of behaviour such as speech, physical skills, and problem-solving are generated in interaction with the environment and not by manipulating previously stored representations in the brain.

4.2.1. An exemplary embodiment position: Andy Clark

In order to explain the adaptive success of creatures we need a new, extended theory with models and adequate mathematical tools. Cognition happens where body, environment and brain are densely coupled. The embodied approaches – and as a representative for them Andy Clark - assume that coping with the world means acting and reacting on continuously changing states of the world. Investigations should focus *on the perceptual as well as on the cognitive contact* with the material world. The prototypical human reason is neither disembodied nor can it be described as consisting of independent modules.[43] Living organisms must be seen as highly integrated systems, which are embedded by means of their bodies in a physical and "effective" (Clark, 1997) environment, or even further in their *Umwelt*[44]. An agent has a body and experiences the world directly via immediate feedback on sensations. The notion of embeddedness encourages us to recognize the role of the body and the fundament-giving environment for cognition. Representations are transformations from an input into an adequate behaviour.[45] As Clark puts it (1997, p.53), the processor as "associative engine" exploits the already existing order and coherence in the environment. The system is decentralised, i.e. it does not have a central executive, no central planner, which directs and coordinates operations. There is no general-purpose inner code, no detailed, action-neutral representation of the world – which leads to the

---

[43] Empirical evidences, based on brain imaging techniques like functional neuroimaging, suggest that complex cognitive acts (e.g. mathematical modelling) result from interactions with the sensory-motor systems. Thus, reasoning results from evolutive pressure and depends on the older phylogenetic structures from which it developed.

[44] Decades ago, von Uexküll introduced the notion of "Umwelt" in another debate. It is the agent's proper phenomenal world, which is also an interpretation of the environment as it fits the agent's capacities and needs.
"We thus unlock the gates that lead to other realms, for all that a subject perceives becomes his perceptual world and all that he does, his effector world. Perceptual and effector worlds together form a close unit, the *Umwelt*" (von Uexküll, 1957, p.6 cited in Ziemke, 2001, p.6).

[45] Here the question arises whether we need a reference to the body only for low-level activities, especially sensori-motor aspects of cognition or if the body should be taken into account for every form of – even high-level – cognition, like reasoning, language-processing, and planning. It seems quite clear that there is a huge difference between a theorem proof and six-legged-locomotion – and such differences are taken into account by the methodological approaches just described. This does not imply that there are completely distinguishable paradigms, but maybe that different cognitive tasks just inhabit different problem spaces, i.e. different extensions of domains. This assumption is the main reason to favour a cluster concept of representation.

slogan that the world serves as its own model. Advanced cognition – not solely low-level processing - results from the interaction of associative engines with their environment. E.g. when one performs multiplication by hand, the process is a collective achievement of brain, body and environmental aspects (ibid, p.61) – cognitive functions are outsourced, off-loaded on the external world. Besides the existence of internal representations, some of the information we use is undoubtedly stored in the environment. Exploiting the structures of the world is called external scaffolding. As the boundaries between perception, action and cognition are vague, the source of cognitive capacities should not be seen as isolated in the skull, but rather as distributed over shifting assemblies including the body and the environmental context. "The true engine of reason (…) is bounded neither by skin nor skull." (ibid, p.69).

Thus, we might consider the mind as subsymbolical bodycontrol[46] of embodied action functioning without a detailed inner model. This leads to a minimal representationalism presented by Clark, which "is a representationalism and computationalism stripped of all excess baggage, and streamlined so as to complement the study of larger organism-environment dynamics" (ibid, p.142). The system does not perceive particular mappings of the world independent from action. Instead it perceives "affordances"[47] for actions having their origin in the close agent-environment coupling. Coupling should be understood here in a way such that the informational content of the representation does not lie only in the inner realms but also in the environment. This leads to the notion of action-oriented representations, which are "representations that simultaneously *de*scribe aspects of the world and *pr*escribe possible actions" (ibid, p.50, my italics). They are local and individual (personal) to a particular agent in its specific situation with the function to guide its actions. To be quite clear, there are representations in this approach - their very existence is not denied, rather there is a reconception. So, what kind of position concerning

---

[46] Bechtel (1993, p.10) gives an example, which shows that the brain may work as an "effective regulator without building up an internal model": Some of the most basic activities of the brain concern the regulation of bodily functions such as the heart rate. For such regulation it apparently does not need to employ a representation of the system it regulates. Of course, it needs some information-source about the system, e.g. the heart, and should be appropriately causally affected, and in addition it needs the means and the procedure to alter the system's behaviour. Bechtel claims that the brain does not have to calculate how the system will react but just let the system respond before itself responding to it. It can adapt to the system by developing appropriate procedures, without representing the external world and calculating responses. And this should hold not only for regulating physiological functions but also for regulating behaviour.

[47] "An affordance is an opportunity for use or interaction which some object or state of affairs presents to a certain kind of agent." (Clark, 1997, p.172) The Gibsonian notion of affordances assumes representations as a collection of possibilities for and restrictions of properties. Cognizers need to generate and make the available information accessible for different uses. An affordance thus denotes the possibilities for actions with reference to the system's capabilities for these actions. Gibson's ecological approach (Gibson, 1979) forwarding a dynamical model of action grounded in affordances, assumes a direct realism opposed to the representationalism of computationalism.

representation-grounding results from this?

4.2.2. Consequences: Is there real connectedness?

The approach tries to ground representational capacities in robotic capacities without a mediation by an external observer. *Appealing to the information from the stimulus via direct perception grounds representations.* There should be an immediate grounding in the real world, which leads to a better representation of it.

Such grounding mechanisms fulfil the central tenets of representation grounding theory: There is meaningful perception, which achieves purposeful action, and finally there are environment dependent symbols. This should not lead us towards an extreme position but rather found a mild shift in our conception.

The thesis of radical embodied cognition[48] completely rejects explanations involving computation and internal representations. I question this extremist position and instead prefer to develop and to improve the computational approach - not to negate it. Such an improvement introduces the concept of "partially programmed solutions", which assumes that there are actually programs[49], which are the basis for further development, formed by evolution. The brain does not have comprehensive a priori instructions for our behaviour, which does *not* deny the very existence of a program at all. Bodily engagement with the objects around us is not specifiable as a determinate representation to which we as subjects have an attitude. How we perceive objects is in first way determined by the interaction of our bodies with these particular objects. We see a body-based form of intentionality here: The body responds to affordances of the situation, without using explicit representations about it. Perceptual states present the world to us in a manner that actually *transcends* our capacity to reflect upon them. In addition**,** Clark (1997) points out that an action is responsive to events in the world which are simultaneously being responsive to our actions – we can imagine an action-interaction cycle, which fits quite well in the previously described alternative approach.

In short: Our fundamental relationship to the outer world needs to be reconceived in such a way that the (human) agent is seen as essentially embedded in a changing world. Still, high-level cognition, as it is mainly studied in Cognitive Science, involves representation. They are not as

---

[48] The Thesis of Radical Embodied Cognition (as it is proposed e.g. by van Gelder, 1990; 1995) summarized by Clark (1997, p.148) is: "Structured, Symbolic, Representational and Computational views of cognition are mistaken. Embodied cognition is best studied using noncomputational and non-representational ideas and explanatory schemes involving e.g. the tools of Dynamical Systems Theory."

[49] Programs are „recipes", i.e. sets of instructions, themselves being a system of signs, which lead to a solution when followed properly. Partial programs deny the total absence of inner instruction sets by allowing a minimal version, which makes the best use out of the given dynamics of the whole system.

basic as assumed traditionally but are derivative from the direct comportment with the world.[50]

5. Conclusions on different levels

5.1. A plea: Representations? Absolutely necessary!

One urgent question remains:

Do we need a reference to the body only for low-level activities, especially sensori-motor aspects of cognition, or should the body be taken into account with highest priority for every form of cognition – even high-level, like reasoning and planning? I argue for the need for representation because especially the high-level cognitive capacities cannot be explained satisfactorily solely by dynamicism and embodiment.

Obviously, representations understood as mere symbols manipulated via inference rules have limited applications in understanding cognition. I agree with scepticism against representations as solely passive, explicit neural instruction sets. In the following paragraphs I will show why we nevertheless need some (extended) concept of representations to investigate and explain cognitive phenomena. It's quite sure in my view, and nevertheless I want to pay some attention to such sceptical positions as risen with the question above. So, let's turn to the evidences.

Take dreams, hallucinations[51], or just reasoning as phenomena, which do not need embedding or dynamics but rather require internal representations for explanation. Dreams and hallucinations clearly demonstrate the capacity of the brain to build complete virtual worlds although there is no sensory input present. It is not self-evident, that this implies that there must be representations, and I'm quite aware that dynamicism might have the means to understand a lot even in this area without representations. I won't go into detail about that in this thesis.

Furthermore, let us consider the causal chain of vision:

There is light from an object in the outer world entering the eye where it is transduced to a neural signal in the optic nerve from where there are subtle transformations to a pattern of activation in the cortex, more explicitly in V1. But it is not only the neurophysiological description, but rather our intuitive view on our world of visual experience indicates that we indeed must possess an

---

[50] This is mainly shown by the idea of bottom-up learning being derived from comportment to conceptual, symbolic representation. In developmental psychology one found out (Thelen and Smith, 1994) that children at first acquire embodied performance abilities, then they learn explicit concepts through so-called representational redescription. Relying on this basis, they further improve performance.

[51] Dreams point on mental activity not ceasing at the onset of sleep. Scientific evidence suggests instead that it is virtually continuous throughout sleep. Hallucinations are false perceptions in the senses, which may appear in hearing, seeing, touching, tasting, and smelling. They are based on no external reality, i.e. in the absence of stimulation. Hallucinations differ from illusions, which are false perceptions based on real stimuli. What we lack when hallucinating, imagining or dreaming is the contextual relation connecting us with the object.

internal map of external spaces, i.e. that we use representations in dealing with the world around us.

In imagery as well as in conception, objects in absence need to be represented somehow.[52] Also, when we hallucinate or dream of a star there is no external cause: Dreams, hallucinations, and visual illusions[53] indicate that the world of experience is not the same as the objective external world that it represents.

A further example giving evidence for our need of representations in the proper study of cognition is the phenomenon of hemi-neglect. It is the lack of attention or even the disappearance of half of the phenomenal space, often as a result of parietal lobe damage. It occurs despite intact sensory and motor functioning. Visual neglect is a common neurological syndrome in which patients fail to acknowledge stimuli toward the side of space opposite to their unilateral lesion.[54] This pathological syndrome obviously also appears to give evidence for an explicit spatial representation in the brain.

One last example from neuropsychology [55]: Take the phenomenon of a retinal after-image, which is seen after the exposure to a very bright light as an internal phenomenon in the eye rather than in the outer world.[56]

5.2. Implications for Cognitive Science

Cognitive Science still relies mainly on the computer metaphor of the mind, but I'm quite sure that this narrow focus alone is not adequate.

Natural cognition serves to organise the agent's adaptation toward viability, which means fitting the environmental constraints with a balance between influences from brain, body and world. Knowledge means knowing how to live in a complex environment in a social community not just

---

[52] The Imagery debate (also called the descriptionist/pictorialist debate) founding on experiments of mental rotation and memorizing maps is able to provide evidence for the very existence of representations in my opinion: Kosslyn (1994), as most important proponent of mental imagery, proposes mental images to represent by virtue of resembling to their objects and by having in fact spatial properties. There seems to be a preservation of topological relationships in the brain, and we seem to be looking at pictures when we imagine things. Kosslyn does not suggest that images are actual pictures in the head but he posits quasi-pictorial representations mimicking a coordinate space. The obvious problem is that we have to import a homunculus looking at the pictures and with that import the problem of an infinite regress. Pylyshyn (1981) argues against the pictorial ideas of representations and states that empirical findings support discursive representations. Tye (1991) proposes a hybrid approach including discursive as well as pictorial elements. In this debate, neither arguments nor counter-arguments are definitive at all.

[54] In the neuropsychological literature it is called "Posner's paradigm." Posner (1980) uncovered a phenomenon in which patients who, while they could see, were unable to process visual information on one side of their field of vision.

[55] For further examples see Gazzaniga et al. (1998).

[56] For a counterexample take a look at van Rooij, Bongers, & Haselager, (2002): They present a typical representation-hungry cognitive problem like imagined action which in fact seems to require representations. But this cognitive task may as well be understood without this concept solely in terms of the Dynamical System Theory.

knowing static facts and procedures that represent what to do.

Models in Cognitive Science must involve representations, they should focus on sensory and motor skills integrating time-aspects, and they should take a decentralised perspective of the neural processes. We need representations because without the idea of some means that intervene between the object and the perception of it, we cannot develop a convincing theory of perception and cognition at all. Yet there is no alternative understanding of "representation-hungry" problems.

We are encouraged to see as complementary the role of the body and the local environment in dense spatial and temporal interaction, including the problem definition as well as the problem solving. The key insights of the different disciplines and different assumptions about how representations mediate between mind and world need to be brought together. So we gain a consideration of cognizers as an overall system - as organisms with their particular environment -, which is compatible with an extended computational and representational approach to the study of cognition.

5.3. Outlook and last words

The notion of computation and mental representation has been rejected in different areas: Anti-representationalist hypotheses have been tested in the psychology of development (Thelen and Smith, 1994), neuroscience (Kelso, 1995), robotics (Brooks, 1991) and philosophy (Peschl and Riegler (1997) and van Gelder, 1995).

In opposition, I suppose that even Brook's approach and the other assumptions from the proponents of anti-representationalism actually need representation at the very end. I'm quite sure that they do need some entity, which explains the ongoing transfer between the inner and the external world.

Indeed, in exploring the field of representations, we are confronted with unanswerable questions[57]. Now - do we really need a revolution in Cognitive Science? I share the view that "minds may be essentially embodied and embedded and still depend crucially on brains which compute and represent"(Clark, 1997, p.143). We have to assume a relocating of the mind: The mind is not understood as an inner entity but as a matter of what we do and how we behave. A resulting invitation might sound: "Take the mind outside the skull" or "Extend the mind" - as Clark and Chalmers (1998) proposed.

Cognition seems to me to be neither solely representational and symbolic nor solely non-

---

[57] E.g. no matter what we think, do, and investigate, we only can check our perception by means of other perceptions, never by the potential objective real entities. E.g. we can never touch, smell, or see the star itself.

representational and dynamical. In the investigation of the mind's access to extra-mental reality, representation should remain a core concept in Cognitive Science, which indeed must incorporate the insight from Dynamical System Theory, embodied attempts, Artificial Life and other alternative approaches to model cognitive processes. Progress needs cooperation: This leads us to a cluster-concept of representation, in which connectionism, Dynamical System Theory as well as embodiment is acknowledged. Such a concept does not require every aspect of it to be realized in each case, but rather gives the possibility to take parts of it and therefore save it from being denied. I argue here for an explanatory liberalism – each strategy can be embraced if it provides helpful means. Even if cluster-concepts might be considered as a compromise in lots of cases, I suggest seeing this particular one as a way to satisfy the criteria given to us when describing and explaining representations and our incredible behavioral capacities.

As it became clear, I'm very sceptic about how we can reach general explanations of much of our adaptive success without assuming information-processing on somehow content-bearing inner states. Denying that representations play an important role in cognition is not convincing at all. Rather, we are in the urgent need of a reconception of the concept. I assume that a well-thought notion of representation which takes an inclusive perspective should maintain truth-conditions because representations have a content that satisfies the external facts or does not.[58]

I agree that it is not informative enough to describe the highly complex inner states as representational in the traditional sense, but this should not be a problem any longer with my extended vision of representation.

Some authors – especially from the embodied cognition research field – warn us not to use the idea of representation automatically and take it as a given as it is done in the symbolists' attempt. To be sure: There is a lot of disagreement and dissatisfaction about the exact nature of the concept. Nevertheless, with my proposed cluster concept I do not assume detailed inner world

[58] Accordingly, one should take care of epistemological questions like whether the world we perceive is the real world itself or merely a perceptual miniature-replica of the external world in an internal representation.
Whereas in the traditional views the relation between knowledge and reality is seen as a (often iconic) correspondence, new approaches describe the relation as an adaptation in a functional sense.
Epistemologically seen, this may lead in the realms of constructivism (see the ideas of Ziemke, 1999 and Peschl & Riegler, 1997): The traditional concept of truth, which indicates a correct representation of reality, is replaced by the idea of viability in the experiential world. Constructivism gives up reality as a way to account for representation and with that looses a necessary and powerful tool of explanation. As there is no place to discuss this issue in length here, I just want to propose the perspective that even strong constructivists like Peschl need some form of representation and in my opinion only shift the problem. In the observer or better in the nervous system of the observer there arise representations – and these need to be explained also by a constructivist account. Gerhard Roth (1997) as a radical constructivist leads himself in a trap, in my perspective, when stating that he has to consider an objective, consciousness-independent or transphenomenal world, called "reality" in contrast to the "Wirklichkeit" as a construct of our brains. The constructor – the brain - must exist in some space and this should be "reality" (ibid, p.288). Thus, he supposes a reality and with that lets representations and the need for an objective reality come into his theory, at least through the back-door. The constructor and the process of construction need a fundament - and in my opinion there is no way out.to assume representations here.

models, which are disembodied and action-neutral, but instead I notice and follow ongoing work in neurosciences. Even with this perspective one should *keep* the idea of internal representations: As involving highly complex, dynamical processes, which are not local. I claim the necessity to rethink the notion of representation in the light of dynamic, recurrent processes, leading us towards a generative paradigm.

„We will see the emergence of new ideas about representation and about computation - ideas that incorporate the economies of action-oriented inner states and continuous analogue processing, and recognize the complex cooperative dance of a variety of internal and external sources of variance.“ (Clark, 1997, p.175).

A further interesting aspect would be to think about questions of priority: Are explicit representations secondary - constructed through bottom-up processes of primordial behaviour and interactive processes? Is there any pre-representational interaction serving for an emergence of representations? Thus, the grounding problem would be solved by assuming that representations are grounded in low-level behaviour.

So as an outlook imagine that when the first step in redefining the notion of representations is to consider the interaction of an organism with the environment, a step further is to consider the coupling not only with a static environment but also with other representing systems. To achieve this, we need an investigation of social and cultural aspects. Social interactions - in particular language – can be studied to learn more about the emergence of representations and "meaning”.

The remaining question:

Why does the brain take all the trouble of constructing a complex internal replica of the external world? Why and how is it done that this replica is presented as such a vivid spatial structure and not as an abstract symbolic code? And of course, how do representations get their content? What is the nature of the standing-in relation?

I suggest that one should not focus any longer on the question of whether representations really exist but instead do good Cognitive Science with investigating what kinds of representations are used in different cognitive processes and how contents of representations may evolve and develop.

6. Literature

Barsalou, L.W. (1999): Perceptual symbol systems. *Behavioral Brain Sciences 22*, p.577-660.

Bechtel, W. (1997): Embodied Connectionism. In D. Johnson and C. E. Erneling, *The future of the cognitive revolution*, pp. 187-208. Oxford: Oxford University Press. http://www.artsci.wustl.edu/~pnp/papers/bechtel.embodied.ps (May 2003)

Bechtel, W. (1998): Representations and Cognitive Explanations: Assessing the Dynamicist's challenge in Cognitive Science. *Cognitive Science 22*, p.295-318.

Bechtel, W. & Abrahamsen, A. (1991): Connectionism and the mind: An introduction to parallel processing in networks. Oxford: Basil Blackwell.

Beer, R. (2000): Dynamical Approaches to Cognitive Science. *Trends in Cognitive Sciences 4*, p. 91-99.

Bickhard, M. H. (1998): Levels of Representationality. *Journal of Experimental and Theoretical Artificial Intelligence 10* (2), 179-215. http://www.lehigh.edu/~mhb0/levels.html (May 2003)

Bickhard, M. H. (2001): Why Children Don't Have to Solve the Frame Problems: Cognitive Representations are not Encodings. *Developmental Review 21,* 224-262. http://www.lehigh.edu/~mhb0/childrenframe.html (May 2003)

Brooks, R. A. (1991a): Intelligence Without Reason. In: *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, p.569-595. San Mateo, CA: Morgan, Kaufmann.

Brooks, R.A. (1991b): Intelligence Without Representation. *Artificial intelligence 47*, p.139-159. http://citeseer.nj.nec.com/brooks91intelligence.html (May 2003)

Brooks, R.A. & Stein, L.A. (1994): Building Brains for Bodies. *Autonomous robots1* p.7-25. Kluwer Academic Publishers, New York. http://www.ai.mit.edu/projects/humanoid-robotics-group/ (May 2003)

Clark, A. (1997): Being there. Putting Brain, Body, and world together again, Cambridge, Mass.: MIT Press.

Clark, A. (1999): An Embodied Cognitive Science. *Trends in Cognitive Science* 3(9).

Clark, A. & Chalmers, D. J. (1998): The Extended Mind. *Analysis* 58 (1), p.7-19. Reprinted in (P. Grim, ed.) *The Philosopher's Annual*, vol XXI, 1998.
http://www.u.arizona.edu/~chalmers/papers/extended.html (May 2003)

Chalmers, D. (1992): Subsymbolic computation and the Chinese Room. In: Dinsmore, J. (ed.): The Symbolicist and Connectionist Paradigms: Closing the gap. Hillsdale, NJ: Lawrence Erlbaum.

Chalmers, D. (1993a): Connectionism and Compositionality: Why Fodor and Pylyshyn were wrong. *Philosophical Psychology 6*, p.305-319.

Chalmers, D. (1993b): A Computational foundation for the study of Cognition. So far unpublished, parts of it have appeared as "On Implementing a Computation" in *Minds and Machines* (1994).
http://www.u.arizona.edu/~chalmers/papers/computation.html (May 2003)

Clancey, W.J. (1991): Situated cognition: Stepping out of representational flatland. *AI Communications—The European Journal on Artificial Intelligence* 4(2/3), p.109-112.

Cummins, R. (1989): Meaning and Mental Representations. Cambridge. Mass.: MIT Press.

Dennett, D. (1991): Real Patterns, *Journal of Philosophy* 88:1, p. 27-51.

Dennett, D. (1993): Evolution, Teleology, Intentionality, (continuing commentary on "Précis of The Intentional Stance"), *Behavioral and Brain Sciences* 16, (2), p. 289-391.

Dretske, F. (unpublished): Experience as Representation.
http://humanities.ucsc.edu/NEH/dretske1.htm (May 2003)

Edelman, G.M. (1987): Neural Darwinism. Basic Books Inc. New York.

Eliasmith, C. (1996): The third contender: a critical examination of the dynamicist theory of cognition, *Philosophical Psychology 9,* 441-463. http://www.arts.uwaterloo.ca/~celiasmi/Papers/thirdcontender.html (May 2003)

Fodor, J.A. (1975): The Language of thought, Cambridge, Mass: Harvard University Press.

Fodor, J.A. (2000): The mind doesn't work that way, The scope and limits of Computational Psychology, Cambridge, MA, MIT Press.

Fodor, J.A. & Pylyshyn, Z.W. (1988): Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition 28*, p.3-71.

Gazzaniga, M.S., Ivry, R.B. & Mangun, G.R. (1998): Cognitive Neuroscience: The Biology of the mind. N.Y., W.W. Norton and Company.

Gibson, J.J. (1979): The ecological approach to visual perception. Boston: Houghton Mifflin.

Grush, R. (1997): The architecture of representation. *Philosophical Psychology* 10, (1).

Harnad, S. (1990): The symbol grounding problem. *Physica D 42,* p.355-347.

Harnad, S. (1993): Problems, Problems: The Frame Problem as a Symptom of the Symbol Grounding Problem. PSYCOLOQUY 4(34) frame-problem.11. http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad93.frameproblem.html (May 2003)

Haselager, W.F.G. (1997): Cognitive science and folk psychology: The right frame of mind. London: Sage.

Haselager, W.F.G., de Groot, A.D., & van Rappard, J.F.H. (2003): Representationalism versus anti-representationalism: a debate for the sake of appearance. *Philosophical Psychology 16*(1).

Hebb, D. O. (1949): The Organization of Behavior: A neuropsychological theory. New York. Wiley.

Heidegger, M. (1927:2001): Sein und Zeit. Tübingen. Niemeyer.

Johnson-Laird, P.N. (1983): Mental Models. Cambridge: Cambridge University Press. Cambridge, Mass.: Harvard University Press.

Kelso, J.A.S. (1995): Dynamic Patterns: The Self-Organization of Brain and Behavior. Cambridge: MIT Press.

Kosslyn, S.M. (1994): Image and brain: The resolution of the imagery debate. Cambridge, MA, MIT Press.

Lakoff, G. (1987): Women, Fire and Dangerous Things. The University of Chicago Press.

Lakoff, G. & Johnson, M. (1999): Philosophy in the Flesh, The Embodied Mind and Its Challenge to Western Thought. Basic Books.

Maturana, H.R. & Varela, F.J. (1980): Autopoiesis and Cognition: The Realization of Living. *Boston Studies in the Philosophy of Science (42),* Dordrecht, Reidel.

McLaughlin, B. & Warfield, T.A. (1994): The Allure of Connectionism Reexamined. *Synthese 101*, p.65-400, Kluwer Academic Publishers.

McLaughlin, B. (1997): Classical Constituents in Smolensky's ICS Architecture, in: M.L. Dalla Chiara et al. (eds.), Structures and Norms in Science, p.331-343, Kluwer Academic Publishers.

Markman, A. B. & Dietrich, E. (1998): In Defense of Representation as Mediation, Psycoloquy: 9, 48 Representation Mediation (1).
http://psycprints.ecs.soton.ac.uk/archive/00000597/ (May 2003)

Merleau-Ponty, M. (1945: 1962): Phenomenology of Perception. Humanities Press.

Millikan, R.G. (1989): Biosemantics. *Journal of Philosophy* 86, p. 281—297.

Newell, A. & Simon, H. (1976): Computer Science as empirical inquiry: symbols and search. *Communications of the Association for Computing Machinery 19*.

Niklasson, L. & van Gelder, T. (1994): Can Connectionist Models Exhibit Non-Classical Structure Sensitivity? In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Hillsdale NJ: Lawrence Erlbaum Associates.

O'Regan, J.K. & Noë, A. (2001): A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24(5).

Peirce, C.S. (1955): The Fixation of Belief. *Philosophical Writings of Peirce* (ed. J. Buchler), Dover Publications, New York.

Peschl, M.F. & Riegler, A. (1997): Does Representation need Reality? Rethinking epistemological Issues in the Light of Developments and Concepts in Cognitive Science. In: Riegler, A., M. Peschl & A. von Stein (1997): Understanding Representations in the Cognitive Sciences. Does representation need reality? Kluwer Academic / Plenum Publishers, New York.

Pitt, D. (2002 Edition): "Mental Representation". *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.).
http://plato.stanford.edu/archives/win2002/entries/mental-representation/ (May 2003)

Port, R. & van Gelder, T. (1995): Mind as Motion: Explorations in the dynamics of cognition. MIT Bradford.

Posner, M.I. (1980): Orienting of attention. *Quarterly Journal of Experimental Psychology* 32, p.3-25.

Putnam, H. (1990): Die Bedeutung von Bedeutung. Ed. and translated by Spohn, W..Frankfurt am Main.

Pylyshyn, Z. (1981): The Imagery Debate: Analog Media versus Tacit Knowledge. *Psychological Review* 88, p.16-45.

Roth, G. (1997): Das Gehirn und seine Wirklichkeit, Kognitive Neurobiologie und ihre philosophischen Konsequenzen. Frankfurt am Main.

Searle, J.R. (1978): Literal Meaning. *Erkenntnis 1, p.207-224.*

Searle, J.R. (1980): Minds, brains and programs. *Behavioural and Brain Sciences 3 (3)*, p.417-457.

Sejnowski, T.J. & Rosenberg, C.R. (1986): NETtalk: A parallel network that learns to read aloud. *Technical Report 86/01*, John Hopkins University Department of Electrical Engineering and Computer Science.

Sharkey, N. & Ziemke, T. (1999): Life, Minds and Robots, The Ins and Outs of Embodied Cognition.
http://citeseer.nj.nec.com/sharkey99life.html (May 2003)

Slezak, P. (2001): Representation: Where Philosophy Goes When It Dies. Proceedings of 23rd Annual Conference of the Cognitive Science Society, Mahwah, N.J.: Erlbaum, p.940-945.
http://www.hcrc.ed.ac.uk/cogsci2001/pdf-files/0940.pdf (May 2003)

Sloman, A. (1995): Towards a general theory of representations. In Peterson, D.M. (ed.): Forms of representation. Intellect Press.

Smolensky, P. (1988): On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11, p.1-73.

Smolensky, P. (1995): Constituent structure and explanation in an integrated connectionist/symbolic architecture. In C. MacDonald & G. MacDonald (Eds.), *Connectionism: Debates on psychological explanation* (p.223-290). Oxford: Basil Blackwell.

Sun, R. (2000): Symbol grounding: A New Look at an Old Idea. *Philosophical Psychology 13 (2)*, p.149-172.
http://www.cecs.missouri.edu/~rsun/sun.PP00.pdf (May 2003)

Thelen, E. & Smith, L. (1994): A dynamic systems approach to the development of cognition and action. Cambridge, MA: MIT Press.

Tye, M. (1991): The Imagery Debate, Cambridge, Mass.: The MIT Press.

van Gelder, T. J. (1990): Why distributed representation is inherently non-symbolic. In: Dorffner, G.(ed): Konnektionismus in Artificial Intelligence und Kognitionsforschung. Berlin, Springer Verlag, p. 58-66.
http://www.arts.unimelb.edu.au/~tgelder/papers/DRNS.html (May 2003)

van Gelder, T.J. (1995): What might cognition be, if not computation? *Journal of Philosophy XCII(7)*, p.345-381.

van Gelder, T. J. (1999): Dynamic approaches to cognition
http://www.arts.unimelb.edu.au/~tgelder/papers/MITDyn.pdf in R. Wilson & F. Keil (ed.), The MIT Encyclopedia of Cognitive Sciences http://mitpress.mit.edu./mitecs, Cambridge MA: MIT Press, p.244-246. (May 2003)

van Rooij, I., Bongers, R.M., & Haselager, W.F.G. (2002): A non-representational approach to imagined action. *Cognitive Science 26*(3), p. 345-375.

Vogt, P. (2002): The physical symbol grounding problem. *Cognitive System Research 3(3)*, p.429-457.
http://citeseer.nj.nec.com/cache/papers/cs/24270/http:zSzzSzwww.cs.unimaas.nlzSzp.vogtzSzcsr.pdf/the-physical-symbol-grounding.pdf (May 2003)

Winograd, T. & Flores, F. (1987): Understanding Computers and Cognition. A New Foundation for Design, New Jersey.

Wittgenstein, L. (1953:1984): Philosophische Untersuchungen. In: Werkausgabe, Bd.1, Frankfurt am Main, Fischer Taschenbuch.

Ziemke, T. (1999): Rethinking grounding. In: Riegler, A., M. Peschl & A. von Stein: Understanding Representations in the Cognitive Sciences: Does representation need reality? Kluwer Academic / Plenum Publishers, New York.

Ziemke, T. (2001): The Construction of `Reality' in the Robot: Constructivist Perspectives on Situated Artificial Intelligence and Adaptive Robotics. In: Foundations of Science Volume 6(1), p.163-233. Kluwer Academic/ Plenum Publishers, New York.