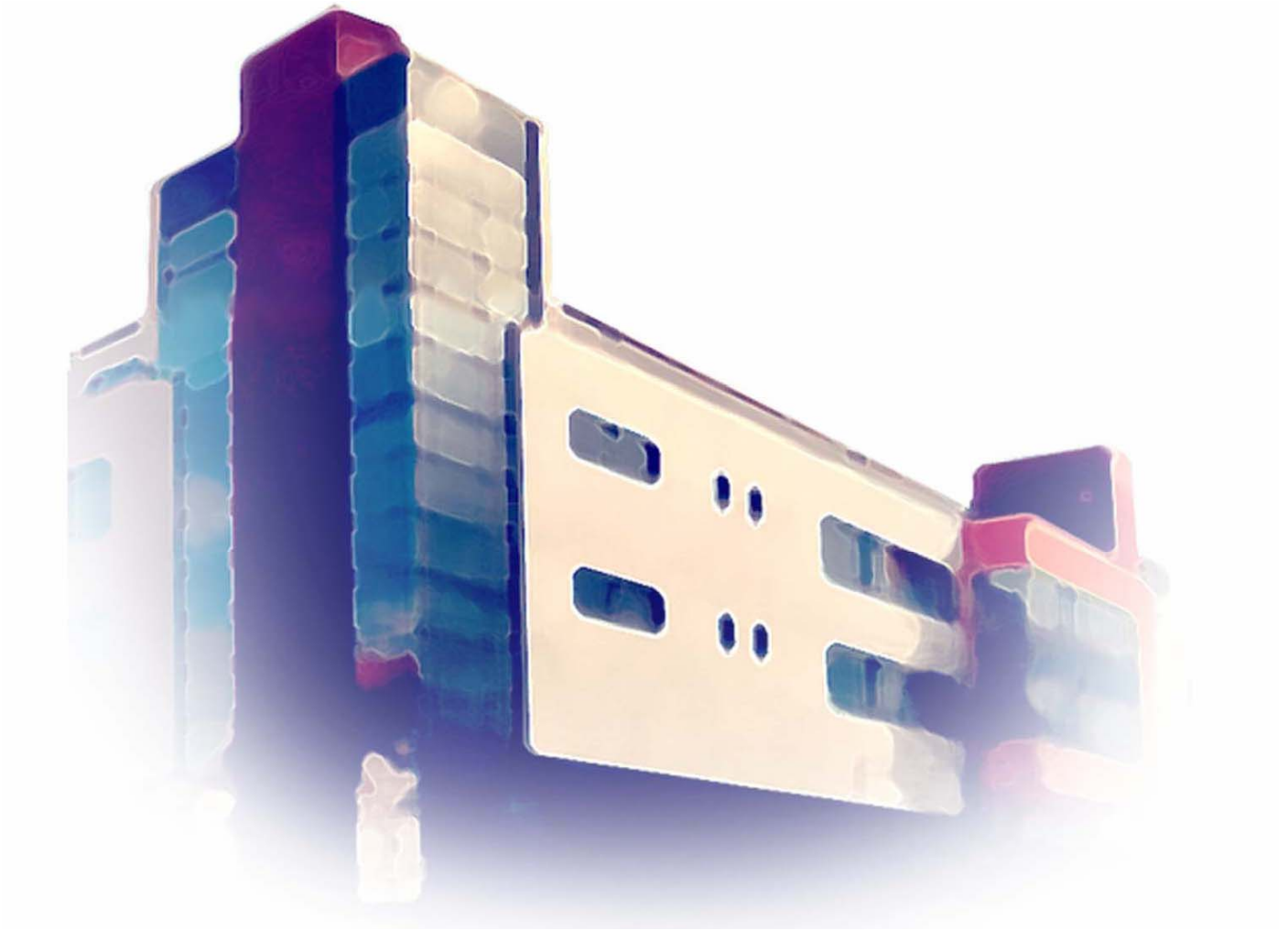
Frank Schumann

Embodied cognitive science: is it part of cognitive science? Analysis within a philosophy of science background



PICS

Publications of the Institute of Cognitive Science

Volume 3-2004

ISSN: 1610-5389

Series title: PICS
Publications of the Institute of Cognitive Science

Volume: 3-2004

Place of publication: Osnabrück, Germany

Date: September 2004

Editors: Kai-Uwe Kühnberger
Peter König
Petra Ludewig

Cover design: Thorsten Hinrichs

Bachelor's Thesis

Embodied cognitive science: is it part of cognitive science? Analysis within a philosophy of science background

Frank Schumann
Cognitive Science, University of Osnabrück
Schumann.Frank@gmx.de
October 2002

Supervisors:
Dr. Ian E. Morley, Senior Lecturer, Department of Psychology, University of Warwick, UK
PD Dr. Achim Stephan, Cognitive Science, University of Osnabrück, Germany

Abstract

The purpose of this paper is to analyse in how far the new field of embodied cognitive science is compatible with the traditional cognitive science program, with a side view on possible differences in the explanatory power between the two different programs.

Two conclusions are drawn. First, the various approaches in embodied cognitive science should be classified into two versions - one of which is compatible with traditional approaches, the second of which is not. In particular, weak embodied theories are still compatible with the core of the traditional program because they still share the latter's computational strategies. By contrast, radical embodied theories are not compatible with the traditional program since they completely abandon the computational approach in favour of the complex interactions between the system, its body and the world.

Second, weak and radical embodiment have different status. Weak embodied cognitive science seems likely to exceed the traditional program in explanatory power, because it can extend the pool of traditional concepts in interesting ways. By contrast, radical embodiment at current seems to be a theoretical position rather than a mature experimental research program, because it is in a process of conceptual and methodological clarification that predates substantial empirical research.

TABLE OF CONTENTS

1 MOTIVATION AND INTRODUCTION

The purpose of this paper is to analyse in how far the new field of embodied cognitive science is compatible with the traditional cognitive science program, with a side view on differences in explanatory power between the two different programs.

The motivation for this analysis is twofold. First, the incorporation of embodied strategies into cognitive science asks for conceptual debate as the altered overall picture of an embodied rather than a disembodied cognitive mind will shape research in important ways. It will not only affect the conceptual language of the program, but also influence what questions are asked, how they are tested, and how the results are interpreted (Beer, 2000, p. 97).

Second, the central question whether embodiment is at all compatible with cognitive science may provide a useful ground in starting such a conceptual debate. In fact, cognitive science fruitfully witnessed a similar discussion previously during the connectionist debate. As in the connectionist debate before, researchers in embodiment now stress that the new approaches differ from the traditional approaches in important ways. And as in the connectionist debate before, at a surface level it is at first is not clear whether the new embodied principles are too different to be part of a cognitive science framework.

I propose the following structure for my analysis. In Section II I will start briefly developing the notion of a research program by Imre Lakatos. I will use Lakatonian research programs to situate my comparison of embodied and traditional cognitive science in a philosophy of science framework of how science progresses over time. My main motivation to choose this strategy is the work of Morley & Hunt (2001) and their course "Philosophy of Psychology" at the University of Warwick. I will indicate Morley and Hunt's reasoning of why to choose Lakatos' work in the philosophy of science rather than that of Popper or Kuhn (who gave other prominent and influential accounts on the nature of scientific progress). In short, Lakatos developed both Popper's and Kuhn's work to include a rational principle on how to actively guide the context of theory discovery. Yet, for those who think philosophy of science is not a normative discipline, this may not seem a necessary step if one wants to compare lines of scientific investigations. But even to those, philosophy of science is helpful in that it descriptively provides a systematic framework for analysis, so that the comparison does not happen at random; and so that the background against which the comparison is made is explicated.

In section III I will first give a general introduction to traditional concepts of cognitive science, and second apply Lakatos's framework to the traditional cognitive science research. Against this, I will compare the embodied research program against later on. The core of the research program of traditional cognitive science shows to be the notion of computation of internal representations.

The next section moves on to discuss embodied approaches to cognitive science. I find it helpful to start with two illustrative examples, and thereafter will give the general and structured investigation of embodied concepts. At last I will again apply Lakatos's framework to what has been discussed. This will identify not one but two research programs of embodied cognitive science. Both stress, but in varying degrees, the importance of environmental cues to intelligent action, resulting in complex interactions between brains and the world.

My emphasis is to find out if the embodied research programs share the computational core of the traditional program in any meaningful way, with a side view on possible differences in explanatory power between the programs.

2 LAKATOS'S RESEARCH PROGRAMS CONSIST OF CORE AND AUXILIARY ELEMENTS

Framework of Analysis

This section motivates a framework of comparison which I will use to contrast the embodied with the classical approaches to cognitive science. This part is mainly based on the work of Morley and Hunt (2001, Ch. 4) and their course on the "Philosophy of Psychology" at the University of Warwick. I will follow Morley and Hunt in taking the view that theories can not be appraised in isolation, and that a rationally developed framework is in urge for the comparison of scientific research. Otherwise comparisons are likely to be arbitrary and at random. I will also follow Morley and Hunt in their reasoning that Imre Lakatos provided an attempt for just such a framework. In short, Lakatos's account assesses conceptual change in science in what he calls scientific research programs. According to Morley and Hunt, in part Lakatos was motivated by his wish to develop a rational element of how to guide theory change after falsification based on two prominent accounts in the philosophy of science: Popper's and Kuhn's. Therefore, Lakatos's own account perhaps is best developed under its preceding influence of both Popper and Kuhn. I will consider each of them in turn.

Popper's principle of theory falsification lacks a rational to structure conceptual change

To start with, it might be said that Popper's philosophy of science was more concerned with falsification of theories than with theories themselves. This means that for Popper philosophy of science focused on truth-statements about theories and not on guiding scientific research in a context of theory discovery. Indeed, Popper considered theory discovery to be psychological rather than philosophical, and thus avoided it. According to Ernest, Popper "explicitly states that in his methodology nothing can be said about the genesis of [theories] and that this belongs to the context of discovery, not the philosophy of science" (Ernest, 1998, pp. 99-100, cited in Morley and Hunt, 2001, Ch. 4, p. 19). This means that for Popper, rational talk about theories is possible only regarding their truth, which can be approximated by the principle of falsification through experimental evidence, but not regarding their development.

This is because Popper's method of falsification itself can not inform theory discovery, since logic retransmits falsity[1] from false conclusions to false premises without indicating which premises were the false ones. If falsification can not indicate which premises are the false ones, but rather equally questions all of them, it can not inform us in any way as to *which* parts of a theory should best be changed in response to falsification. Put in another wording, this means that Popper could not establish a connection from modified theories to

[1] Logic is useful because valid logical arguments deduce true conclusions out of true premises. But, logic also transmits falsity backward from false conclusions to false premises. The argument runs like this (Hunt, 1999). Forward truth-transmission is a property of logic's syntactic structure that premises (implicitly) entail all true conclusions. Then, however, the fact that true premises in valid arguments can only lead to true conclusions also means that false conclusions require false premises. "Otherwise the 'true' premises would entail a false conclusion: a contradiction of the notion of validity." (Hunt, 1999, p. 14). This principle of *retransmission of falsehood* means that false conclusions must be asserted to at least on false premise. Unfortunately, however, which of the premises is/are false cannot be concluded from a false conclusion. Therefore, falsification merely shows that one or more of the premises must be false, but not which. A serious limitation of the principle of falsification.

their precursor theories. Even though scientists factually do change theories at least in part to respond to falsification, Popper's philosophy of science has to treat any changed theory as *completely new* theory even if the theory is an improved old one.

A second aspect of what this lacking connection between old and new theories means is that Popper's work on scientific change misses out on elements that would allow to *structure* scientific progress over time, and that Poppers methodology can not guide theory discovery in scientific processes. Without a rational to structure science, however, comparison and appraisal of potentially competing theories such as traditional and embodied cognitive science appears to be difficult or arbitrary.

In contrast to Popper, Kuhn's paradigms do structure conceptual change, but more for social than for rational reasons

Kuhn's philosophy of science is also well know and influences many empirical researchers that are not themselves philosophers of science. At first sight, his account seems splendid to compare research programs, because contrary to Popper Kuhn explicitly aimed for a structure of scientific progress in what he called *scientific paradigms*. Such paradigms are defined as "periods in which certain kinds of methods and theories are sufficiently dominant to define … science within that period" (Morley and Hunt, 2001, Ch. 4, p. 15). However, Morley and Hunt provide at least two major problems with Kuhn that eventually lead them to support a Lakatonian view in the philosophy of science. First, the dominating methods defining a Kuhnian paradigm are not chosen rationally, but socially. This means that for Kuhn, the paradigm-defining methods and concepts are just those to which a group of scientists, for whatever reason, happens to agree. As Morley and Hunt put it, in Kuhn "the main reason for paradigm choice is *social*, rather than *intellectual* [and thus to some extend irrational]" (ibid, p. 15). In a descriptive reading this implies that Kuhn assumes science to be a social rather than a rational process. And in a normative reading then also Kuhn's account cannot provide the rational reasons to guide scientific progress that where already lacking in Popper.

The second problem with Kuhn is the common critique "that the notion of a paradigm as applied to psychology … sees too little diversity within … the same paradigm (or schools or research traditions)" (p. 18).

If Morley and Hunt are correct on both issues, the Kuhnian account to structure scientific progress is a) not fine-grained enough, and b) still lacks a rational other than social convention. Then, comparison of research fields in a Kuhnian manner again seems to be difficult and in danger of being arbitrary.

Lakatos's term 'research program' seeks a rational to structure scientific change in the light of its historical development

Lakatos's account can be understood as an direct attempt to improve both Kuhn's and Popper's ideas. With respect to Kuhn, Lakatos seeks to avoid Kuhn's overly social criteria in structuring science. He seeks rational criteria that structure science in the historical development of research fields. With respect to Popper, Lakatos holds on to the principle of falsification, but additionally aims to find the principles to guide theoretical responses to theory falsification in the context of theory discovery, that were lacking in Popper. I will consider both aspects in turn, and end this section with how I will apply Lakatos framework in the next parts of the paper.

Kuhnian periods of dominating paradigms are criticised because they are defined purely by social rather than rational criteria. Lakatos takes up this work of Kuhn on the social influences on science, but develops it to avoid an *overly* social bias. He placed "falsification in a wider context of rational choice, showing that responses to falsification made *sense* in their *historical* contexts" (in Morley and Hunt, 2001, Ch. 4, p. 18, my emphasise). Accordingly, he structured science considering historical developments of a line of scientific change, in what he called *scientific research programs*. A research program shapes the conceptual background for the day to day activities of researcher. As Winograd and Flores put it "[t]his background invisibly changes [and thus shapes] what they choose to do and how they choose to do it". (1987, p. 25). Lakatonian *research programs* are constituted by fundamental *core assumptions* that all theories within a program have to share, and by *auxiliary theories* that implement this core. In this structure, "everyday" research within a program typically occurs on the auxiliary level; it involves finding and adjusting the right auxiliary hypotheses to further specify the core ideas, but everyday research typically leaves the core itself untouched. Changing the core on the other hand is more radical and may modify the fundamentals of the research tradition[2]. Lakatos regards scientific development that changes to much of the core of a program not as occurring inside, but outside of that program. That is, radical changes of the core create whole new programs rather than mere adjustments within a program.

Such changes in the core most closely mirror Kuhn's notion of shifting research paradigms. However, Lakatos's more elaborated concept of core and auxiliary elements enables more fine-grained distinctions between various types of research than Kuhn's paradigms. Also, Lakatos is said to use the term "research program rather than [Kuhnian] paradigms to emphasise the *active* role that a research program plays in *guiding* the activity of scientists" (Winograd and Flores, 1987, p. 24, my emphasise).

Missing out on active guiding principles is a critique also commonly associated with Popper. Popper has nothing to say on how a scientist should respond after falsification. This is because falsification equally questions *all* of the premises of the falsified theory, and Popper did not distinguish between core and auxiliary elements of a research program. As a result, in a strict reading of Poppers philosophy of science, falsification of a single theory *always* brings the whole research program into question. For Lakatos, by contrast, the normal response to falsification happens as an adjustment of auxiliary elements, and does not affect the core of a program. The core, so to speak, entails the concepts that researchers would "change last", whereas the auxiliary levels contain elements that researchers would "change first" (Ian E. Morley, personal communication). The auxiliary levels thus are said to form a *protective belt* that preserves the core from refutation. The *negative heuristic* is Lakatos first guiding principle for scientific change after falsification: progress (and daily research) within a program means to find the right auxiliary hypotheses that implement the program's core by conducting a sequence of experiments and falsifications. This leads to a connected sequence of more and more improved theories that specify in more detail what is meant with the core of the program. However, in daily research, the core itself remains untouched and is assumed as a set of background assumptions.

---

[2] However, *resorting* the importance of individual elements in a program's core may not yet call out for a radically new program. This is relevant, because weak embodiment eventually will resort some elements of traditional cognitive science in their importance.

This way of guiding scientific change would of course be open to "dogmatism" (Morley and Hunt, 2001, Ch. 4, p. 19) and thus itself irrational, if Lakatos had not added a second heuristic principle, the *positive heuristic,* in prevention. It is that changing a theory should not simply block its falsification[3]. Rather, changed theories should also gain in explanatory power. Research programs which loose explanatory power are said to *degenerate*, and are likely to be replaced by new and more powerful programs. Only programs that gain in explanatory power are said to *progress*.

Two more aspects of Lakatos's work gain importance. First, Lakatos's view opens choice, rather than empirical necessity as an option to decide in favour of two research programs. Lakatos considered the core of a research program as being a *metaphysical core*, because he thought that the core was not open to empirical testing. This is partly because "very few theses are testable in isolation", and thus empirical testing involves "many other assumptions and theories to render [theories] testable" (Morley and Hunt, 2001, Ch. 4, p. 21). Because part of the core's function however precisely is to provide the theories with which to test the auxiliary theories, we would need yet another set of ideas against which to test the metaphysical core. Consequently, the core of a program cannot be tested by direct empirical testing. From this point of view, the only empirical criterion to decide between two competing programs is to wait until one of them degenerates.

Second, Lakatos distinguishes between "'mature science', consisting of grown up research programmes, and 'immature science', consisting of 'a mere patched up pattern of trial and error'" (Winograd and Flores, 1987, p. 24). Part of what this means is that in the early periods of a new research program, researchers have to spend substantial time not only on doing experiments, but also on developing an adequate conceptual language (including mathematical and formal tools) needed as a basis for experimental work (John, 1998, p.7). In John's view, this has a profound implication, namely that "experimental results become relevant only after a certain 'warm-up period' of the program" (ibid, my translation). As an example, Bechtel (1998) argues that such a starting period was granted to connectionism after its second birth. These last two arguments will become important in my attempts to asses the status of weak and radical embodied research programs at the end of this thesis.

Conclusion

To compare traditional and embodied cognitive science, I will choose Lakatos's term research program because it provides a structural component that is lacking in Popper, and avoids the social bias in Kuhnian paradigms. However, within the scope of this work, Lakatos terminology can not be applied in all its complexity, but only in some detail. Thus, in the following I will confine myself to distinguish research programs in three basic levels that are implicit in the discussion on Lakatos above.

- First, elements on the ***metaphysical level*** are the ones that define the program. These are the concepts researchers would change last.
- Second, the core is further implemented or interpreted by important concepts and methods on a first auxiliary or ***implementational level***.

[3] Lakatos also calls this the *positive heuristic,* whereas he calls the core's protection by auxiliary assumptions the *negative heuristic.*

- Third, a yet more auxiliary or ***practical level*** contains more practical implications that follow out of the first two levels, such as the specific kinds of problems studied in the daily research of the field.

This perspective is powerful enough to allow a graded perspective on scientific change, in that changes on the outer levels do not affect the nature of the program as much as do changes on the inner level. Additionally, this perspective sets out criteria for appraisal of radical scientific changes: to be a true follower of the Freudian research program, one has to share the metaphysical core of Freudian theory. If one violates this core, one can no longer call herself a true Freudian and enters (or creates) a new research program. To end with, during their formative years Lakatos's perspective grants new programs a 'warm-up' period, which they need to develop their own conceptual language.

3 TRADITIONAL COGNITIVE SCIENCE

Framework of analysis

Cognitive Science differs from behaviourist, cultural or social traditions in psychology in at least two respects. First, it seeks to understand aspects of intelligent action as made possible by a system's internal capacities, rather than aspects made possible by outside factors. And second, internal capacities of a system in turn are thought of as a complex information-processing device that "receives, stores, retrieves, transforms and transmits information" (Stillings et al, 1998, p. 1). Early researchers saw information processing analysis of the mind as so powerful that they did not only believe that some form of it was both necessary and also sufficient to understand the mind, but also believed that "[within] a generation the problem of creating 'artificial intelligence' will be substantially solved" (Minsky, 1967, p. 2, cited in Dreyfus, 1992, p. xlvi).

Part I of this sections aims to show the motivation for such (over-)promising statements. Taking up Lakatos's spirit that theories can not be praised in isolation, I think this can best be understood in front of the historical background grounding the field. Part II of this section then aims to identify the structure of the traditional program in cognitive science in a systematic way, so that the structure of the embodied program can be compared to it later on. The term traditional cognitive science as used here refers to the first period of cognitive science (Varela, Thompson and Rosch, 1991) that had symbolic artificial intelligence at its centre. The term traditional does not refer to connectionist strategies, and these will only briefly be mentioned where it seems relevant, but not discussed in particular detail.

3.1 Introduction to traditional cognitive science

In sum, the term cognitive science subsumes various computational theories of mental phenomena. Their computational nature is what unifies the multiple disciplines in the field and may count for much of its success in recent years: computational theories combine the analytical tradition of formal enquiry with mechanical implementation of such theories in computers. Establishing this combination was powerful, as true AI for the first time in human history in principle seemed achievable. Therefore, a satisfactory introduction to cognitive science should be twofold. First, it should discuss the analytic tradition of scientific thought, with some emphasise on how mechanical computation can give causal power to analytic theories. And second, it should then go on to discuss the application of such computational theories to cognition. I will consider both aspects in turn.

Computational theories of mind stem out of the analytic tradition of enquiry, which seeks syntactic and context-free theories. Those can be mechanised, giving causal power to logical form

Traditional cognitive science stems out of the analytic tradition of inquiry, which in general gains broad appeal as scientific method in the western culture. It should be looked at

here for two reasons. One, to appraise cognitive science it is helpful to understand the "taken-for-granted background shaped by the underlying assumptions of the rationalistic tradition" (Winograd and Flores, 1987, p. 26). And second, such analytical analysis produces formal theories which, if sufficiently detailed, can be implemented on machines. This gives analytical theories causal power in real world, and inspired the hope for some form of artificial intelligence.

What, then, is the analytic tradition? For Winograd and Flores, analyzing a phenomenon means to:

> "1. Characterize the situation in terms of identifiable objects with well-defined [that is, identifiable] properties;
> 2. Find general rules that apply to situations in terms of those objects and properties;
> 3. Apply the rules logically to the situation of concern, drawing conclusions about what should be done " (Winograd and Flores, 1987, p. 14ff, cited in Hunt, 1999, p. 4ff)

Hunt adds a second perspective on the analytic tradition given by Dreyfus&Dreyfus:

> "The success of theory in the natural sciences … reinforced the idea that in any orderly domain [of enquiry] there must be some set of context-free elements and some abstract relations among them that account for the order of that domain and man's ability to act intelligently in it [i.e. to investigate and understand the domain]" (Dreyfus&Dreyfus, 1988, cited in Hunt, 1999, p. 4).

According to Hunt, the key element of the analytic tradition described by both characterizations is its "formalist, symbolic and calculative" core (Hunt, 1999, p. 5). This means that analytic inquiry seeks to separate the phenomenon under study into atomar elements, and then applies formal rules to describe the symbolic elements and the relations between them (ibid, p. 4). This strategy seems promising because "what is assumed in the analytic tradition is that by symbolising elements and relations, these descriptive strings of symbols (symbolic sentences) can be logically transformed [that is calculated] into derivative strings of symbols which explicitly answer the questions asked in the domain" (ibid).

Two properties of formal theories open them for mechanical implementation: they are syntactic, and they are context free. First and perhaps most fundamental to formal analysis is the belief that the syntactic properties of some phenomenon can be formally described by syntactic properties. The form of some elements is important in formal theories, not what the elements actually mean or stand for. Morley and Hunts put this nicely in a strong view on syntax and semantics which means "roughly […] that if you look after the syntax the semantics will look after itself" (2001, Ch. 15, p. 12). Second, the analytic tradition seeks to employ only context-free elements in its theories. This is necessary, because formally valid transformations can only preserve truth when the semantics of a theory does not depend on some context (Hunt, 1999, p. 3). That is, when the semantics only depends on the elements and the relations that are entailed in the theory, and not on anything else.

Then, what is most striking about formal models in the context of cognitive science is that they can be mechanised - executed by machines "without intuition or interpretation" (Dreyfus, 1992, p. xi). This result out of two seminal developments in mathematics and logic (Gardner, 1985). Alan Turing found a theoretical machine that can compute any mathemati-

cal function. Later, von Neumann build a mechanical implementation of Turing's theoretical machine with his General Purpose Computer.

Mechanical implementation of cognitive theories seems promising for two reasons. First, again, it gives causal power to logical form, as emphasised for example by Fodor (e.g., 2000, p. 19). And second, mechanical implementation renders the implemented theories open for empirical testing. Placing strong emphasises on the first point leads to the position of 'strong AI' or 'cognitivism', which believes that it is in principle possible to create artificial systems that have genuine cognitive states and do not only mimic them.[4] Placing the emphasize on the second point leads to a softer position of 'weak AI', in which AI is appraised only as a powerful methodological *improvement* in the study intelligence, rather than the ground laying framework in which genuinely intelligent artificially systems will be possible[5]. What is seen as improvement is that theories must be sufficiently specified to be implementable on machines. Thus, computational theories of the mind are said to pose much harder constrains on theoretical accuracy than before in psychology (Searle, 1980, in Beckermann, 1999, p. 277)[6].

Symbols and representations: physical symbol systems as necessary and sufficient means for intelligence

Computational theories of course are interesting only if they are applied to something. Here, Newell and Simon perhaps most strongly influenced the application of computational theories to the field of cognition, combining formal analysis of the mind with artificial creation in machines. In their hypothesis, a *physical symbol system* (PSS) that certifies certain conditions provides the necessary and sufficient means for human intelligence (Newell and Simon, 1972, in Morley and Hunt, 2001, Ch. 14, p. 22).

The idea is that intelligent systems are physical systems (be it brains or von Neumann computers) in which some physical structures serve as symbols that represent some parts of the world for the system. The claim then is that the behaviour of such a system is determined by the symbol structures it implements and by the processes that operate on those symbols, but not by the system's particular physical structure itself. That is, the claim is that any physical system that implements the same symbol system will show the same behaviour. The central thesis of traditional cognitive science then follows quite naturally. It claims that to analyse how a PSS performs intelligent thinking, it is sufficient to study these symbols and

---

[4] The most famous critique of this strong view on AI is perhaps Searle's Chinese room argument, which argues that syntax is not sufficient for semantic. If syntax does not yield semantics, then formal systems can not have intentional states, because by definition they only operate on syntax (Searle, 1980, in Beckermann, 1999, p. 279). However, Searl's argument is a thought argument, and it likely might be one that belongs into the class of bad thought arguments.

[5] AI of course is also of tremendouse use for engineering purpose outside of psychology. These, however, are not in the focus of this paper.

[6] However, also this weaker form of AI is not without its critique, as large parts of those harder constraints posed by computational theories of mind reflect merely unimportant details of technical implementation rather than genuine properties of cognitive functions. It is thus disputable whether more precicion automatically gives better theories. That is, the precision of a computational over a psychological theory on itself not necessarily offers more insight into cognition. This leads to an even weaker appraisal of AI, in which AI theories offer value for the study of the mind only by means of predicitons that are verified in simulation experiments. For an argument that even these simulated experiments have little explanatory relevance if they occur in simulated "toy worlds", as they commonly do, see the work of the Dreyfuß brothers.

how they are computed, because the exact physical structure of the system does not matter in creating intelligence apart from implementing a particular symbol system (Beckermann, 1999. p. 265ff). This is important. It rendered symbolic analysis the fundamental heuristic of classical research in cognitive science. To quote Howard Gardner:

> "the triumph of cognitivism has been to place talk of representations on essentially equal footing with … the neuronal level … [or] the socio-cultural level" (Gardner, 1985, p. 383).

Unfortunately, the term symbol often conflates two meanings (Chalmers, 1992) which should be kept separate. First, symbols serve as representations. But second, the term symbol is often used to specify those representations that have syntactic format. Symbols function as representations because they "stand-in" for objects or events in the world so that the system can operate on the representation instead of the objects directly (Bechtel, 1998, p. 3).[7] This gives important advantages. For one, a representational system can use environmental features even if they are not "reliably present to the system" (Haugeland, 1991, in Clark, 1997, p. 144). Also, if reasoning processes operate on representations rather than on the world directly, the system can generate a variety of different options before it manipulates the world in ways that can possibly not be undone. Indeed, one important feature of rational thinking is the ability to contemplate about and choose between different options before one actually changes the world by some action (Winograd and Flores, 1987, p. 97).

The second aspect of the term symbol then specifies the format of such symbolic representations. Traditional cognitive science demands that representations have symbolic, that is, syntactic format. If representations have syntactic format, they can serve as elements in formal theories that describe how new representations can be produced by the combination of other representations according to syntactic computational rules. Symbolic mental representations are thus the most direct way to combine syntactic computation with the kind of analytic analysis described above..[8]

What has been said gives a general motivation for computational analysis of intelligence, and as a consequence it also gives a particular motivation for Newell&Simon's physical symbol hypothesis. Both motivations stem out of the powers gained by the mechanisation of analytic theories. Yet, two further assumptions on the nature of the human mind made by Newell and Simon are essential to illuminate their approach. In their first assumption they views flexibility as the central feature of the intelligent mind. In fact, Newell and Simon stem both the necessity and the sufficiency claim of their physical symbol hypothesis out of the

---

[7] Philosophers have taken several approaches as to what exactly it is for a structure to "stand in" for something else and thus be called a representation; and certainly not any kind of internal structure serves as a representation (Clark, 1997, p. 147). Too touch just two aspects of the discussion, representations can either be defined in terms of the objects they stand for, or in terms of the processes that use the representation, or in terms of both (Bechtel, 1998, p. 3ff). Dretske for instance defines representations by the idea that they carry information over the object they represent. That is, for him a state is a representation if there is another process that uses the state in virtue of the information it carries. On the other hand, Milikan stresses that representations have a certain function for some process. Then, a representation is defined by the function it plays in and for some process, and not necessarily by the information it carries about an object. Bechtel thus concludes that any account on what a representation is must involve three aspects: the thing represented, the representation, and the process that uses the representation.

[8] A weaker formulation demands representations only to have *some* structured format which enable structure sensitive processes to operate with them, but not necessarily syntactic format (Clark, 1997, p. 145; Stillings et al., 1995, p. 344). This weaker formulation then allows connectionist structures or other types of encoding schema that may be used by the mind or in the brain to serve as representations.

astonishing flexibility they see central to human *and* computer intelligence, stressing that this flexibility can only be gained by symbol systems (Morley and Hunt, 2001, Ch. 12, p. 29) that are not restricted by the particular physical implementation. This is the necessity claim that any system capable of extreme flexibility will turn out to be a physical symbol system, because a symbol system's "ability […] to maintain internal representations of the world, to access them and to transform them with processes that are not driven by sensory input or tied directly to motor output […] is crucial to its flexibility" (Stillings et al., 1998, p. 24). The sufficiency claim is that any physical system which is a symbol system can be developed such that it exhibits this flexibility (Morley and Hunt, 2001, p. 24). This reasoning views input and output structures as constrains that limit a system's flexibility, the goal then being to overcome these limitations. To anticipate, part of the embodiment movement is to recognize that input-output structures not only constrain but also *empower* a system because they are adapted to the world such as to exploit specific environmental cues for concrete (rather than general) forms of intelligent action (Clark, 1997, p. 87, in Morley and Hunt, 2001, Ch. 15. p. 13).

The second (closely related) assumption on the nature of the human mind that is implicitly hidden in the physical symbol hypothesis is that *abstract rational processes* are the distinctive feature of the intelligent human mind, as opposed to creative, tacit[9], or emotional[10] processes. To quote Stillings et al., the "physical symbol hypothesis is motivated in part by the human capacities for planning, problem solving, and reasoning …" (1998, p. 24). As a result, rational forms of thinking such as reasoning and decision making in abstract domains compose the heart of research in traditional cognitive science.

At last I think it is worth noting a reformulation of how physical symbol systems operate given by Andy Clark. He thinks that due to the van Neumann style of computation contained in physical symbol systems "we are talking about the kinds of basic operations provided in a general purpose digital computer supporting some high level programming language, such as USP [or] PROLOG …" (Clark, 1989, p. 12, cited in Morley and Hunt, 2001, Ch. 12, p. 12). Any reader with experience in such programming languages as PROLOG will perhaps easily understand why Marvin Minsky had to reformulate his appraisal of the AI problem: after 15 years of research he considers it to be "one of the hardest science has ever undertaken" (cited in Dreyfus, 1992, p. xi).

Cognition and intelligence conceptualized as problem solving

The discussion above showed how belief in the rational nature of the human mind fuels the attempts to formalize mind as physical symbol systems. But Newell and Simon went one step further in conceptualizing what they think the mind does. In their seminal book "Human problem solving" they equal rational thinking with some form of rational problem solving (Newell and Simon, 1972). The reasoning is that because intelligent minds show abilities for rational problem solving in a variety of domains, they might simply be devices for problem solving in *general*, which constantly apply their general problem solving abilities to dif-

[9] From a philosophical point of view, Dreyfus (1992) for example bases his critique on the possibility of strong artificial intelligence on stressing the importance of tacit knowledge in human thinking. See also the work of Daniel Denett and Gilbert Ryle. From a psychological perspective, see Claxton (2000) for an account of non-rational processes in human thinking.
[10] Neuropsychological evidence for the importance of emotional processes in general and in rational thinking in particular is given for instance by Damasio (1996).

ferent domains. Then, so the reasoning, most of human cognition can be studied as just one instance of general problem solving.

Very broadly construed, such general problem solving involves several steps. In the first step, the system symbolically represents the current state of the task environment, the desired goal state and the available actions (called operators) that can change the state of the system. In the second step the system then searches through the search space of possible actions for the ones that lead to the goal. In the last step needs to decide between multiple actions that could fulfil the goal for the ones that are fastest, or bring the highest value at the lowest cost. This strategy also entails a view on decision making in which decision making means to predict he utility of all possible outcomes (that are available to the system) and to choose the one with the highest value (Winograd and Flores, 1987, p. 20ff).

Winograd and Flores think that this view of cognition as problem solving is a view "generally taken for granted in artificial intelligence research" (Winograd and Flores, 1987, p. 22). Indeed, not only AI but also research in cognitive psychology is largely influenced by this view. Anderson, for example, still structures his introductory textbook on cognitive psychology in major parts around the belief "that all cognitive activities are fundamentally problem solving in nature" (Anderson, 2000, p. 240).

As a result of the above conceptualizations, the input-output picture of the mind stresses the independence of cognition from sensory input and motoric output systems

A last broader picture of the relation between the mind and its environment can be formulated, which underlies the traditional attempts to formalize the mind. As discussed, these attempts are motivated in part by the assumed independence of rational intelligence from sensory-motor processes. As a direct consequence that is inherent in this spirit, cognitive science generally neglected generation of input and execution of output. In fact, only a slide reformulation of general problem solving described above uncovers what Susan Hurely (1998) calls the 'input-output picture of the mind'. First, a *perceptual* systems takes a snapshot of the world, classify and conceptualize the world in explicit, symbolic representations. Second, the *cognitive* engine uses the logical rules of inference-machines to compute some output on the basis of those representations. And third, the cognitive system finally *instructs* the output devices; again using explicit, symbolic representations. Susan Hurley's point is that here the description of a traditional cognitive system would end. It typically does not include how perceptual input to the cognitive system was generated in the first place, nor how the output of the cognitive system is finally executed by the motor systems. Perception, cognition, and motoric execution remain distinct. Perceptual input is *already presented* to the cognitive mind, whereas output of the cognitive mind only *instructs* behaviour. Already the symbolic instructions for actions are considered as the output of the cognitive mind, not the actual actions themselves. This is to say that this "input-output picture" sums up to a clear-cut separation between mind (internal computation) on the one hand, body and world (input generation and output execution) on the other hand.

Conclusion

Part I showed the roots of cognitive science in the analytical tradition of philosophical enquiry. Because such formal theories are syntactic and context-free, machines can do the rule transformations in such theories. Machine implementation then gives causal power to logical form, and renders computational theories testable. Newell and Simon see one form of computational systems as necessary and sufficient means for human intelligence: physical symbol systems. The behaviour of PSS is determined already by the symbolic structures, whereas the physical structures that implement those symbol structures don not themselves add anything in. Thus, system behaviour can be understood at the representational level. Newell and Simon ground their hypothesis in assumptions about the human mind: its flexibility and its rational nature. Furthermore, they conceptualize cognition yet more precisely as instance of general problem solving. In sum, the input-output picture of the mind that results out of this line of research stresses the independence of cognition from sensory input and motoric output.

3.2 The research program of traditional cognitive science

This part analyses the structure of the traditional research program in cognitive science, so that the embodied research programs can be compared against it later on. To recapitulate, I consider a research program as defined in what Lakatos calls the 'level of the metaphysical core'. Any theory that aims to be part of cognitive science has to share this level. More auxiliary concepts that further implement this core are contained in a second 'implementational level'. A last 'practical level' provides place for rather practical implications that follow out of the first two levels. For instance, such as the specific problems researchers select in their daily research activity.

The metaphysical core: the computational metaphor of the mind

To start with, the unifying view that all theories in cognitive science will have to share is that of information processing in terms of computation over internal representations. Analysis of a system's internal representation of the world and computational theories of how the system processes that knowledge give causally plausible models of the human mind. These are, so to speak, the ingredients that traditional cognitive scientists would "fear to loose" (Ian E. Morley, personal communication).

Entailed in this view are systemic boundaries between mind and world. For a system to act on internal representations requires that those representations are distinct from the external world, otherwise the systems mind would perceive and act in the world *directly* (Clancey, 1997).

In its early days, representations in cognitive science had to have symbolic format, and computation was equalled with rule based, syntactic manipulation of those representations. However, as Lakatos stresses that theories should not be praised in isolation, a quick look at the connectionist debate might be helpful in appraising the traditional AI program. Connectionism started a dispute on whether *symbolic syntax manipulation* is at the core of cognitive science. If that were the case, it would exclude connectionism from cognitive science. How-

ever, I think today the common sense view on the connectionist debate (e.g. Clark, 1989) is that connectionism is part of cognitive science, providing a first useful supplement of concepts and strategies to the field. It allowed non-symbolic representations in cognitive theories, and thus shifted symbolism from its early place at the very core of the traditional program to the implementational level. Then, after the debate the core of computation over internal representations could be further implemented either by symbolic or by non-symbolic means. Therefore questions on the nature of representations shifted from the core of cognitive science to the implementational level. However, what is still and importantly left as the core of cognitive science after the connectionist debate is the notion of computation over internal representations.

The implementational level: the mind as problem solver

The common core of internal computation of representations is further spelled out and constrained at the implementational level of the traditional research program in cognitive science. First, as just seen, in the traditional view the notion of computation over internal representations is equated with symbolic information processing. Second, the symbolic information processing strategies that underlie cognition in turn are further restricted to search and planning strategies. The computational metaphor is narrowed down to the mind as **problem solving system** in the sense of AI.

The methodological tools of traditional cognitive science then are those of problem solvers as know in AI:

- symbolic knowledge representations languages;
- problem solving and logical inference
- the use of algorithms for search and planning;
- regarding the body only as input/output device

The practical level: abstract high-level and offline reasoning

In my analysis, what has been said so far entails many implications that I consider to be more of practical nature, shaping the surface level of the traditional program. My argument later is that these will be challenged most by embodied theories. And as we shall see the practical level in embodied programs looks much different.

- Because only internal maps of the world are thought to be responsible for guiding the systems behaviour, *all* behaviour has to be computed explicitly.
- Therefore, the world has to be represented as detailed as possible to cover all features of the world that could *possibly* be relevant for action. Clark calls this strategy "representation hungry", involving rich, or "heavy representations" (Clark, 1997).
- Heavy representations entail large demands on memory and computation time, and would therefore ask for off-line reasoning, computing the full action sequence before action starts, rather than online reasoning, computing the next step while already acting.

These practical considerations also contribute to problem domains that are studied in traditional cognitive science. These are for instance:

- The study of human abilities in abstract logical reasoning in cognitive psychology
- Formal mechanisms of language processing, such as automatas and artificial grammars in computational linguistics
- The study of formalisms for knowledge representation, such as propositional logic, mathematical feature vectors, scripts, schemas, knowledge languages, frames, semantic networks, mental models and so on both in AI (see e.g. Russel and Norwig, 1995) and cognitive psychology (see e.g. Eysenck and Keane, 2000). All of which are thought to underlie particular forms of human knowledge but not others.
- The study of problem solving strategies, involving questions on search (e.g. heuristic search in AI; or the use heuristics in of human reasoning in cognitive psychology); machine learning in AI (e.g. the ID3 algorithm); memory models, allocation of processing time, and formation of concepts in cognitive psychology (see e.g. Eysenck and Keane, 2000)
- In general, avoidance of problems that involve timely delicate interaction with the environment because of the off-line reasoning strategies that result out of the large demands on computing time,
- Rather, most research involves high-level tasks (for instance planning systems such as chess programs, or expert systems like MYCIN in medicine), rather than lower level tasks (such as perceptual systems for instance).

Finally, emphasise on internal representation was accompanied by "de-emphasise on affect, culture, and history" (Gardner, 1985, p. 41). It is substantial to my argument to point out that Gardner thinks this de-emphasise was more due to practical consideration, rather than to principal ones. Therefore, if Gardener is correct, de-emphasise of the outside is not part of the cognitive science's metaphysical core. And consequently, emphasise of the outside does not necessarily affect the core of the program. Therefore, embodied theories have good chances to remain compatible with cognitive science.

4 EMBODIED COGNITIVE SCIENCE

Framework of analysis

Natural environments and physical bodies provide important valuable clues for intelligent action. Yet, in the last section perceptual- and motor systems were considered as a restriction rather than an empowerment of intelligent action. By contrast, embodied cognitive science aims to exploit these environmental cues by studying how evolutionary developed, domain-adapted cognitive systems cope with their natural environments. To do so, new concepts are being developed. Some of these new concepts may stand in competition to the traditional view of internally computed instructions for action. But some of the new concepts may only alter or amend traditional tools. Whether embodiment is compatible with traditional cognitive science depends on how radical these new concepts differ from the old, and on what place they will gain within the new program.

The following analysis starts with two examples of embodied research. Studies on autonomous robots will illustrate the idea of scaffolding and the active use of the environment in problem solving. Studies on childhood locomotion development stress the complex web of causal factors with which the mind is related to its body and to the world. Part II provides and discusses some more embodied concepts, such as emergence, structural coupling and dynamical systems theory, using a hierarchy of four claims in embodied theories put forward by Andy Clark. Part III then sorts what has been discussed into the framework of Lakatos, contrasting the embodiment research program with the traditional research program of the last section. The main source of this section is the work of Andy Clark (mainly 1997 and 1998).

4.1 Introductory examples: how to exploit the environment to reduce computational cost

(Example 1) Autonomous agents: collecting soft drink cans at MIT

Autonomous agents are embodied systems that navigate themselves through complex natural, rather than much simpler simulated environments. Parts of the demands that natural environments pose on such agents is the ability to act quickly, to rely on incomplete, noisy and contradictory information, and to adopt to new and different environmental settings. NASA funded robots such as DANTE II, for instance, autonomously have to explore the largely unknown surface of planets. The embodiment community commonly justifies its endeavour with stating that traditional symbolic AI performed only poorly on such tasks.

One of the earliest autonomous robots is Herbert (in Clark, 1998). MIT designed him to collect empty soft drink cans in the MIT robot laboratory. To do so, he had to navigate autonomously through the lab, and not bump into moving people or other robots.

Herbert's designers wanted to avoid shortcomings of traditional systems and used "the simplifications and shortcuts afforded by simple environmental cues" (Clark, 1998, p. 508).

The traditional approach would have resulted in planning systems enhanced by sensory modules such as image scanning and image processing capabilities, which would build detailed and computationally demanding online maps of the laboratory. But, if a robot was to live in a real rather than a toy environment, why not directly use as much information given by that environment as possible without first modelling it, thus reducing computational cost and complexity?

Herbert reduces computational cost because he is build according to what is called *subsumption architecture*. Many different subsystems independently trigger different aspects of the robots overall behaviour without central control. These systems can make use of parts of the environment, and at the same time neglect elements not relevant for their tasks, thereby reducing the overall computational cost. In Herbert's case, all those different systems were very simple. He had only a simple motor system to move him, guided by ultrasonic sensors that could stop movement when obstacles were immediately near by. A simple visual system could do no more than detect outlines of tables; a second video camera could scan for the outline of cans. Even the arm collecting the cans moved blindly, aided only by simple touch sensors that skimmed the surface of the table. If Herbert found such can-like shapes, a simple grasping behaviour was activated to get the cans.

Herbert illustrates some of the features that are attributed to embedded autonomous agents. First, to navigate he did not have nor build and update detailed internal models of the laboratory. Second, he did not have to plan or infer anything on the basis of those models. That is to say that Herbert was not guided by detailed sequences of actions based on an internal map of the lab, but by simple cues out of the environment such as 'obstacle ahead' that trigger equally simple actions such as 'move backward'. In this sense, Herbert's perceptions (e.g. 'obstacle ahead') are said to make 'immediate sense' to him (e.g. move backward).

To the embodiment community, such illustrations seem convincing. But I think there at least two things that need further discussion. First, although Herbert employs a subsumption architecture that is opposed to central processing mechanisms, one might argue that such central mechanisms are implemented purely by the design of Herbert's hardware. He doesn't need flexible, non-hardwired central processing because his abilities are highly limited. He moves without knowing where he goes, he grasps without knowing what he grasps. He can not make decisions but *must* grasp anything that is sufficiently similar to a can on a table, regardless if it is one or not, since he has almost no conceptual and categorizational abilities. To some thus it might ask for much credit in advance if they are to appraise the subsumption architecture approach as an achievement with an example as limited as Herbert. Yet, perhaps because of that Herbert serves as good illustration of the general idea of the embodiment movement: *Herbert's limitations build into his body are to his favour*. Traditional planning systems fail remarkably in real-world navigation tasks not despite but because they aim for high conceptual and categorizational abilities to build detailed inner maps from which to infer relevant action sequences. The point made by the embodiment thesis here just is that much of everyday action does not need the flexibility gained by high-level centralized planning. Rather, systems that rely on many limited, but evolutionary developed domain-adapted and distributed mechanisms can manage autonomous navigation within environmental niches much better than traditional general planning systems. From an evolutionary perspective, such an approach gains much plausibility. Evolution will develop "quick and dirty" solutions to very specific demands, rather than general problem solving strategies.

To many (traditional) criticizers of embodiment, the second point to note follows straight out of the above, and concerns the viability of an (entirely) embodied program: can high-level reasoning systems be build without central planning systems? The feature of Herbert's abilities was that they are highly limited, and it is not clear how systems exclusively built by such methods might develop abstract high-level reasoning. The obvious way out of this high-level problem is that the use of domain-depended subsystems in some domains (such as navigation) does not exclude the option to build inner representations in other domains. In more complex systems, embodied principles possibly are connected to internal maps at different levels. Another way to resolve this is on the line that abstract reasoning skills in fact do develop out of concrete reasoning skills, e.g. in the spirit of Piaget or Dennett.[11]

(Example 2) Childhood Development: How the situated infant learns to walk

Child development provides my second route into embodied cognitive science. Clark (1998, 1997) reports Thelen and Smith's (1994) work on the "stepping motion reflex" of infants. How do infants know when to perform their first stepping motions? New-borns, if "held suspended off the ground, will perform a recognisable stepping motion. After a few months, this response disappears, only to reappear at about 8-10 months … [and] independent walking cuts in at about a year" (Clark, 1998, p. 506). But what guides the development of locomotion in infants? One type of answers involve a 'grand plan', or a single factor that governs the onset of walking. There might possibly be a program for genetically "predetermined phases of […] locomotion" (ibid). In the same manner, simple development of neural locomotion patterns in the brain or in the spine might cause the infant to start stepping.

By contrast, Thelen and Smith promote a somewhat different, embodied type of answer: They see the *complex interactions* of multiple factors that connect the body, the brain and the world as the triggers for locomotion development. Their multi-factor view gains support by two different external factors that can trigger the stepping reflex. First, holding the baby upright in water provokes walking even in periods where stepping motions are normally not seen. And second, letting the baby step on a treadmill has the same effect. Thelen and Smith report that two different parameters trigger the walking in the two cases – bodily and external factors, rather than internal ones. In the first case of holding the baby upright in water, they key parameter seems to be simple leg mass. The infant refrains from stepping in certain periods simply because the mass of the infant's legs prevents him from doing so, and not because of any internal program. In the second case of the treadmill, "a kind of mechanical patterning caused by the backward stretching of the legs" (ibid) is responsible for the onset of walking. In conclusion, Thelen and Smith think that the development of infant motion is due to "the interplay of a variety of forces spread across brain, body, and world", rather than due to any internal program. Such forces "include bodily features (such as leg mass), mechanical effects (stretch-and-spring), ecological influences (water, treadmill), and cognitive impetus (the will to move)" (ibid).

In sum, both examples illustrate the common ground of embodiment perspective: what seemed to be complex cognitive phenomena requiring the development of complex reason-

[11] The program of Lakoff and Johnson (e.g. 1986) in cognitive linguistics for instance is an explicit version of this kind of thought, in which they pose that all human concepts, even abstract ones have concrete concepts gained by experiences of the body as their underpinnings. However, I will not discuss them in detail here, as their work is based on language analysis and does not particularly attempt to inform cognitive architectures (Raffael Nuenez, personal communication).

ing or planning capabilities turns out to be performed by simple hardwired mechanisms that are distributed all over the systems body and triggered by environmental cues.

4.2 Introduction to embodied cognitive science

The discussions in the embodiment literature of course involve many more issues than the two examples could provide. In the following, I discuss more of these further issues using a structure provided by Andy Clark. He thinks that the various theses on embodiment fall within one of four claims that summarize embodied cognitive science. The first two of which are compatible with the traditional picture of cognitive science, the latter two of which are not.

To anticipate, in more recent work Andy Clark himself supports a weaker version of embodied theories and thinks of the last two claims as being too radical. The 1999 article clearly is critical to radical views on emergent explanations and dynamical systems theory. However, in most parts of the earlier 'Being there' (1996) Clarks own position remains considerably unclear until towards the end of the book.

4.2.1 Weak embodiment: Clark's claim one and two

Clark summarises the common core of the weak version of the embodied approach in two claims:

> **Claim 1: "That attention to the roles of body and world can often transform our image of both the problems and the solution spaces for biological cognition" (Clark, 1998, p. 506)**

Much of what this claim involves is the extension of problem solving routines with *local, action-oriented representations*, and *epistemic actions*.

Local representations and external scaffolding

Locally effective representations save much of the computational power needed for detailed categorization of objects because in given environments much simpler cues of the environment with less demands on categorizational processing and memory can sufficiently identify an object for a given task. Herbert demonstrates some of these new principles. For instance, Herbert's representations of soft drink cans are only locally effective, saving computational cost at the price of generality and flexibility. All Herbert can notice are can like shapes above a table'. He can't explicitly identify those shapes as soft drink cans or even differentiate between cans and other objects of similar form. Would MIT's researchers use bottles instead of cans, Herbert's strategy would miserably fail the new demands. The reduction of computational load is made possible by a restriction to only locally effective representation, confining the discriminatory efforts only to what is needed for the given task in the given environment.

A second example is Ballard's (1991, in Clark, 1997) work on animate-vision. In a similar manner, the visual system represents only locally effective features of the environment. For

instance, once you know that your cup of coffee is yellow, you don't need to identify your cup *as* a cup each time you reach for it. You rather can simply pick it up by reaching for the yellow object next to you. But again, this strategy is limited to that particular environmental setting. For Ballard, what might seem like a limitation actually shows to be a common feature of our cognitive system. Ballard thinks that we do not only heavily use local environmental cues for cost efficient representations, but that we "*actively structure* [our] environments in ways that will reduce subsequent computational loads" (Clark, 1997, p. 150). This is possible because a system only needs to represent "those aspects that are relevant for guiding behaviour" (Bechtel, 1998, p. 3). In a broader sense such an active use of the external structure is what developmental psychologist Vygotski meant by the term *scaffolding*, which he used to describe how external structure informs the development of an individuals intrinsic structure. The idea is that individuals develop within a zone of proximal development that is given by social and other external context. That is, individuals need and are bound to some external input to develop the intrinsic structure.

Action-oriented representations

Action-oriented representations are a second strategy to save computational cost. They *simultaneously* represent the world and appropriate actions (relative to the given environment), in that they already include "a specification [of appropriate] motor activity" in representing an object (Clark, 1997, p. 151). Herbert again represents the environmental cues on which he relies in a strictly action-oriented manner. If he senses an obstacle in front, all he can do is move backward. If he senses a can, all he can do is to grasp. He hasn't much choice. In contrast to traditional planning systems, his input representations are already directed toward more or less specific actions. Therefore, the system does not have to infer informational value and appropriate actions out of detailed world models in complicated problem solving routines. Rather it can avoid such heavy demands on memory and computation time, because the "early encoding are already gathered toward the production of appropriate actions" (Clark, 1997, 152). Those encodings do already have meaning in terms of a variety of possibilities for actions even without further inferences. The concept of action-oriented representations is similar to Gibson terminology of perception as *affordances* for action (Clark, 1997, p. 172). The system does not perceive action-neutral mappings of the world, but a large variety of possible *behaviours*: affordances for action provided by the systems close coupling to the environment.

Here we find a first way to understand the coupling of a system to its environment. Local and action-oriented representations are valuable only in an appropriate environment. Part of the informational value of the internal representations lies directly in the external world, which means that it is lost if that environment changes.

Epistemic actions

Acting can occur for two reasons: to execute a priory solved solution to a given problem, or as part of the problem solving process itself. Acting in the world to find a solution for a problem is another shortcut out of the embodiment toolbox that can drastically save computational cost and memory. Using such *epistemic actions* (Kirsh and Maglio's, 1995, in Clark, 1997), Tetris players for instance rotate the blocks *on the screen* and in the actual game, and

not exclusively in their heads. Rather than solving problems on internal models before any acting starts, embodied systems might act *as part of the problem solving process itself*.

In conclusion, the above principles are less computationally expensive supplements to traditional problem solving routines. They are computationally efficient, because they reduce the representational complexity so that the agent has to solve less representation hungry problems (Clark, 1997). By doing so, they also support a form of *online reasoning* by acting in the world, rather than *offline reasoning* on internal models. However, in contrast to general problem solving routines, these strategies are only efficient in specific domains: they couple the agent to the environment. For Clark, the above concepts represent the central ideas in embodiment: "a cognitive science that takes [these concepts] seriously … will have gone a long way towards remedying [the traditional] disembodied intellectualist bias" (1998, p. 511) and thus have moved towards embodiment.

For the purpose of contrasting research programs, let us be clear about what is *not* being said here. What is not being said here is that local and action-oriented representations were *in*compatible with computational and representational stories of the mind. They do extend the repertoire of traditional problem solving routines with computationally less demanding strategies, adding a new meaning to the term representation. But we may be confronted with supplements to the problem solving mechanisms of the traditional program, not with principle arguments against it.

Emergent behaviour, self-organization and structural coupling

Clark's second claim is:

> **Claim 2: that "understanding the complex and temporally rich interplay of body, brain, and world requires some new concepts, tools and methods [for] the study of emergent, decentralised, self-organising phenomena"(Clark, 1998, p. 511)**

In complex systems of many elements, we often find some behaviour that is not centrally controlled or explicitly programmed; it rather 'emerges' out of the sheer interactions between the individual elements of the system. An example will help to illustrate. Resnick (in Clark, 1997, p. 110ff) programmed artificial ants to collect and sort wood chips into piles. To start with, in a classical non-emergent solution to this problem a central control program would give instructions to each single termite on which chip to take up and on which pile to leave it. Resnick's emergent solution, by contrast, doesn't need such a central program to command the ants. Rather, Resnick's ants act autonomously according to two very simple rules. First, if an ant does not carry anything and bumps into wood, it shall pick up the wood. Second, if it bumps into wood when it is already carrying wood, it shall leave its carriage at that place. What is important is that both rules directly instruct only *individual ants*, but do not directly instruct to the ant colony as a whole. And still, acting in accordance with those two rules made the ant colony sort 2000 wood blocks into 34 piles. The point to be made here is that although *the colonies* behaviour *is* fully caused by the interaction of the individual ants,

do the two rules that instruct *single* ant behaviour not yield a full explanation of *colony* behaviour.

Indeed it is at first not clear why Resnick's strategy works at all. It first allows removing wood from piles as easily as adding it. Second, it does not provide a terminating signal stopping the ants from infinite shifting of piles to different locations. But still the program terminates. The trick is that once all wood is removed from one location, that location is effectively blocked. This is because the two rules do not allow creation of new piles at blank locations. And therefore, the number of piles must decrease over time. This blocking feature however was not explicitly programmed anywhere in the program. It rather emerged from the interaction between the rules (which don't allow new piles) and the environment (which provides only a limited amount of wood piles). It is a feature of the whole system, and not of individual ants. The whole seems to be more than the sum of its parts.

The example illustrates in an oversimplified way what is meant by emerging properties. More generally, emergence is very close to the notion of self-organization. Systems show self organizing behaviour if on a higher level patterns emerge out of the interaction of lower-level components "without the benefit of a leader, controller, or orchestrator" (Clark, 1997, p. 73). Rather, these systems self-organize themselves. First, in a relation of 'upward causation' (Thompson and Varela, 2001, p. 418), the behaviour of the parts causes some particular behaviour of the system. And second, in a relation of 'downward causation' (ibid), this particular overall behaviour of the system feeds back to the action of the parts.

But let us again be clear about what is proposed here. The terms self-organization and emergence often seem to be used to express 'unexpected' behaviour of a system one cannot readily explain; and the usage of the terms seems to conflate different meanings at different times. However, first, emergent explanations used in this sense[12] are fully reductive explanations, in that the high-level behaviour is fully caused by low-level components without any mystical processes involved. Even the properties Varela and Thompson call to be caused by downward causation are in fact fully caused by the individual parts, even if the downwardly caused properties can only be instantiated when the individual parts play together. Even if the individual parts themselves are changed by the downward processes, this is fully redundant if in principle it is understandable a) how the individual parts kicked of the higher processes, and b) how the downward processes change the individual parts. A first reading of 'emergent explanations' then just refers to the absence of central control mechanism (Clark, 1997, p. 110).

Second, although emergence is often applied to structurally coupled systems, the term only may but does not have to entail a notion of structural coupling. Clark (1997, p. 73ff) distinguishes both options. Direct emergence describes processes in which some overall behaviour directly emerges out of internal elements of the system *alone.* Indirect emergence, by contrast, requires that the interactions between the systems elements are *mediated* by environmental structures. It is this second reading emphasizing some form of structural coupling that seems to inspire much of the use of the term in embodied cognitive science.

[12] For a stronger, nonreductive conception of emergence see Stephan (1999).

Third, structural couplings in turn may or may not involve central control mechanisms. Again, the term emergence is often used to emphasise structural coupling even when it is not clear whether central control mechanism are at hand or not.

Fourth, the term structural coupling itself conflates several meanings and can be differentiated on the size of systems it is applied to. Ziemke (2001) stresses that widely read it is not of much use to cognitive science, because "every system is in one sense or another structurally coupled with its environment" (Riegler, in press, cited in Ziemke, 2001, p. 3). Ziemke thus attempts to classify five notions of structural coupling used in embodiment that increase in restrictiveness; however a detailed discussion of those is not substantial for the present purpose.

Dynamical systems theory as the methodological tool for emerging behaviour

Emergent systems often include many variables. Dynamical systems[13] theory is a common tool to study emergent and self-organizing behaviour because it can produce complex and unexpected behaviour out of only a few variables. DST "reduces potentially very complex [high dimensional explanations] to simpler, more tractable [low-dimensional] ones" (Clark, 1998, p. 512). By contrast, traditional attempts to model complex behaviour quickly involve vast amounts of variables and become intractable. Connectionist models for instance often require hundreds of parameters to represent input and output neurons. Therefore, recently DST is used more and more often in a variety of domains, including low-level phenomena such as motor coordination as well as high-level phenomena such as decision making (for a detailed overview see Port and van Gelder, 1995; for a short overview see e.g. Jaeger, 1995). Especially theoretical neuroscientists heavily use DST, e.g. to model the chemical and spatial properties of individual neurones (e.g. Dyan and Abbott, 2001).

4.2.2 Radical embodiment: Clark's claim three and four

The use of dynamical systems theory as a replacement rather than a supplement to traditional explanatory tools in the cognitive science toolbox leads to the last two and most radical of Clark's claims. It is that

> **Claim 3: once the new concepts are in place, "these new concepts, tools, and methods will perhaps displace (not simply augment) the old explanatory tools of computational and representational analysis" (Clark, 1998, p. 512)**

Dynamical systems theory may abandon representational analysis of the mind

A common reasoning by advocates of dynamical systems theories is that it abandons a representational story of the mind. The most obvious reasons would be their premise that DST itself is non-representational (van Gelder and Port, 1995). If this was correct, DST and representational analysis would be mutually exclusive. However, the premise that DST is non-representational is subject to discussion. DST could be judged as non representational, first, because it does not involve the discrete symbolic states common in the classical notion of a representation. The states in dynamical systems rather continuously change over time.

---

[13] An understandable introduction to dynamical systems theory and its application to cognition is Jaeger (1995). For a comprehensive discussion see for instance the Behavioural and Brain Sciences articles of van Gelder's (1998) and Thelen et al. (2001).

Second, DST might not be representational because of the different, limited kind of explanations it yields. It (only) "highlights a global behaviour at the expense of functional detail" (Clark, 1998, p. 513). Restated in Bechtel's (1998) words, DST does not yield 'mechanical explanations' of how the system's internal parts produce a certain behaviour. Rather, the DST approach yields (only) 'law-like explanations' *describing* the system's behaviour. That is to say that dynamical systems only name a rule which describes how the behaviour of the system looks like, but not how it is generated. Indeed, the parameters in such a rule do "not [even] have to correspond to components of a system which interact causally. They can be, rather, features in the phenomenon itself" (Bechtel, 1998, p. 17).

Interestingly, if correct, the view that DST and representational analysis are mutually exclusive supports not one, but two competing conclusions. One reading is the above rejection of the computational metaphor. In this line of thought, if DST for whatever reason should give the more adequate kind of explanation for cognition, but does not yield the required mechanical descriptions of discrete internal states and subcomponents, then the representational story seems to be mistaken. In a second reading, however, we could alternatively question DST as explanatory principle and reject Clark's claim three. In this line of thought, what we expect of a theory of mind are mechanical explanations about the internals of a system, and how exactly those internal factors are influenced by external factors. If DST explanations can't offer us both of these, we might thus reject the dynamical systems approach as the solitary way to study cognition. Some seem to have taken the former view (e.g. van Gelder 1995, Port and van Gelder, 1995); others have taken the later (e.g. Clark, 1997).

William Bechtel (1998) rejects the mutual exclusiveness premise of above. He attempts solve this dispute between DST and representational analysis by arguing that DST indeed can be of representational nature. In his view, DST extends the term representation in interesting ways, rather than to replace it. He argues that there are structures in dynamical systems that can have what he sees to be the fundamental characteristic of representations: to "stand in" (ibid, p. 3) for something else. He recognizes that dynamical strategies *do* differ in two aspects from traditional representational strategies, however, he thinks that the differences are interesting ones and amount to new concepts that (nicely) augment, and not dispel representational analysis. First, representations in dynamical systems differ from classical representations in their *format*. Most notably, representations in DST are dynamical rather than static. Bechtel argues that representations are made up of three components: "what is represented, the representation, and the [processes operating on] the representation" (Bechtel, 1998, p. 5). He agrees that in contrast to symbolic representations, the dynamical representations of DST can not be identified as representation by some processes that operate *on* them. However, dynamical structures such as attractors and trajectories still can *serve* as representations *in* some processes, as do connectionist representations. Bechtel argues that the criteria that render neural networks representational also render dynamical systems representational. If Bechtel is correct, than DST adds a new format of dynamically changing representations to the symbolic and connectionist formats already in use. And, there seems to be a need for such dynamically changing representations, for instance in neuroscience.

The second difference was already noted above: dynamical systems differ from traditional representational strategies in the type of explanations they yield. Dynamical explanations are "law-like' description of the system's behaviour rather than detailed 'mechanical' explanations of how the internal of the system produces that behaviour. For Bechtel, traditional mechanical explanations seem compelling at the *intra-systemic* level that seeks to identify the

working of the systems subcomponents. By contrast, law-like representations can be useful especially at the *inter-systemic* level on which one studies how several systems interact, rather then the individual systems themselves. This is interesting, because at the inter-systemic level a good description of "what [a] system is doing" might be sufficient even if one does not know "how it does it" (ibid, p. 19).

Bechtel sums up in that, "[t]he larger point to be made […] is that cognitive science has explored a wide variety of representational formats. DST, by introducing new notions such a trajectories and dynamic attractors, contributes to this ongoing exploration" (ibid, p. 11). Clark arrives at a similar conclusion, in that we will "need a mix of levels of analysis [… and] explanatory tools, combining Dynamical Systems constructs with ideas about representation [and] computation" (Clark, 1997, p. 123).

There are, however, some other reasons why dynamical systems likely will not replace traditional representational analysis, at least not in the near future. Two major challenges for today's dynamical systems are their limited practical applicability and the narrow scope of phenomena that might be studied using them. At first, the mathematics of DST becomes more and more intractable with increasing numbers of parameters and dimensions of the system's state space. In fact, the majority of differential equations used in dynamical systems there have no known analytical solution. Which means they can only be studied by numerical methods. Second, in his commentary on Van Gelder's (1998) BBS paper on "The dynamical hypothesis in cognitive science", Herbert Jaeger argues that cognitive systems have characteristics that even the dynamical models of today can not handle. First, DST principles like attractors and bifurcation are not of much help in wild systems with "fast stochastic input" (ibid, p. 643) varying on the systems own characteristic time-scale. Second, as already stressed above, DST handles high-dimensional domains by reducing them to low-dimensional descriptions. In line with Clark and Bechtel above, Jaeger takes the view that this reduction to some collective parameters is helpful in some respects, but still poses a limit to the study of wild, high-dimensional systems. Third, Jaeger thinks that 'wild' cognitive systems are non-stationary, meaning that there is some nonparametric change in the dynamic law itself. And "with non-stationarity of the strong kind […] we are simply lost" (ibid).

However, Jaeger thinks that these limitations may eventually be overcome with further development of dynamical systems theory. In particular, he argues that 'hybrid models' combining dynamical with symbolic approaches will bring major advances in studying cognition.[14] Yet, hybrid models to often are more easily posed than developed.

Continuous reciprocal causation may remove systemic boundaries, and render analytic analysis impossible

Clark's fourth claim is:

> **Claim 4: "That the familiar distinctions between perception, cognition, and action, and indeed, between mind, body and world, may themselves need to be rethought and possibly abandoned" (Clark, 1998, p. 513)**

[14] His own 'observable operator models' (OOM) shall provide a start for such hybrid models. http://www.ais.fraunhofer.de/INDY/herbert/Publications.html, September 2002.

This is the most provocative, radical and ambiguous claim Clark ascribes to parts of the embodiment discussion. Adopting to this claim means to abandon the separation between the world and the mind that set up the systemic boundaries between a system and its world - a radical and highly counterintuitive proposal. In particular, the claim is that if the causes for intelligent behaviour were equally distributed between the internal of a system and the world, then this would rob talk of internal structure of its meaning. Instead, the feedback loops between internal and external factors rather than the factors themselves are then seen as the cause of intelligent behaviour.

However, we should appraise this claim with caution and clearly distinguish principle reasons that would substantially support the claim from practical reasons for which the claim might only seem obvious. The sheer complexity of the causal interactions might indeed very well cause *methodological* difficulties in separating between internal and external causal forces. Thelen and Smith speak of an "inextricable web of perception, action and cognition" (Thelen and Smith, 1994, p. xxii, in Clark, 1998). However practical problems by no means imply that the mentioned inextricability is a principle one. And therefore practical considerations alone do not yield the metaphysical conclusion given in claim fourth. What would be needed are principle reasons.

Such principle reasons to abandon systemic boundaries might be given by considerations on circular causation emphasised for instance by Thompson and Varela (2001) and Varela, Thompson and Rosch (1991). In what Clark calls 'continuous reciprocal causation' (1998, p. 514) Varela et al. present a form of circular causation in which "no discrete temporal staging interrupts the coupled dynamics of linked systems" (ibid), because the relations between the systems are analogue and not digital. In such systems we perhaps still can identify single components of the systems, but we can not explain the individual behaviour of the components by isolating them. Clark thinks that "given the continuous nature of the mutually modulatory influences, the usual analytic strategy of divide and conquer yields scant rewards" (ibid, p. 514). This means that looking at individual units can not even give us their own behaviours, because they are not the sufficient objects of study even in terms of *their own* behaviour. Rather, the necessary object of study then would be the larger system the elements are part of, even if we are only interested in the behaviour of individual components. Clark concludes that if the causal relations between mind, body and world are indeed continuous and reciprocal, "we may indeed confront behavioural unfoldings that resist explanation in terms of inputs to and output from some supposedly insulated cognitive engine" (1998, p. 515). That is to say that separating the individual system from the global system would not be possible.

In conclusion, claim four faces serious scepticism. First, DST is, at least at present, not powerful enough to serve as the only explanatory tool. And second may the reasons to accept claim 4 mainly be practical, but not principle ones. However, embodied strategies might become handy in covering these practical challenges. Especially in cases where law-like descriptions of the systems behaviour as given by DST can already be useful for some purpose, even if a detailed mechanistic explanation is not yet possible.

4.3 The research programs of embodied cognitive science

As done in the section on traditional cognitive science before, Part III of this section now analyses the structure of embodied research programs in cognitive science, contrasting it to the traditional program. I will, however, apply Lakatos's structure of a research program only to what I call the weak version of embodiment. To recapitulate once more, a program is defined in what Lakatos calls the level of the metaphysical core, an implementational level with further specifications of the core, and a practical level with practical implications out of the first two levels. Radical embodied research alters the core of traditional programs. But as will become apparent, I think that radical embodiment is not yet sufficiently developed to apply Lakatos's structure in detail.

4.3.1 Weak embodied research programs are compatible with computational cognitive science

The metaphysical core is compatible with the computational metaphor of the mind

I propose that weak embodiment shares the metaphysical core of cognitive science outlined in section II: information processing in terms of computation over internal representations. A first reason to do so is Andy Clark's appraisal that even though weak embodiment does stress the role of the environment in intelligent action, it considers the environment as "just a source of input to the real thinking system, the brain" (Clark, 1997, p. 105) . This remains compatible with a computational theory of this brain. A second reason for suggesting that embodied ideas do not touch the metaphysical core in radical ways is Gardner's remark at the end of section III. He thinks hat dis-embodiment was rather due to practical reasons and is not an explicit part of the metaphysical core of traditional cognitive science.

But where did the focus on disembodied theories then come from? I suggest that the neglection of the body results from a one-sided reading of Newell and Simon's physical symbol hypothesis. Recall that according to the hypothesis the power of the symbolic mind lies in its independence from its sensory-motor systems. It does then not come as a surprise that research at first focused on disembodied, sensory-motor independent strategies. But Newell and Simon themselves posed the PSH as an empirical hypothesis. It must show to be sufficient by the extend to which it empirically shows useful in creating artificial intelligence, and in so far as psychological experiments will reveal a symbol-manipulating nature of human thought (Newell and Simon, 1976). Then, following the embodiment argument, the independence gained by the PHS empirically showed to be useful mainly in high-level reasoning tasks, but not (or to a much lesser extend) for low-level tasks involved in direct action with real worlds. Hence, the shift of interest towards embodied action lead to new embodied strategies that aim to use domain-dependent low-level strategies to exploit environmental cues.

However, this shift in interest does not imply that we have to abandon the old computational mechanisms at all and forever. Rather, we may extend the old tools by embodied strategies. Similarly, Markman and Dietrich (2000) conclude in their review on alternatives to classical representations

> "that representation should remain a core part of cognitive science, but […] the insights form these alternative [embodied] approaches must be incorporated into models of cognitive processing" (Markman and Dietrich, 2000, p. 471).

On the implementational level, the mind as controller of embodied action uses a combination of embodied and disembodied strategies

I propose that incorporating embodied approaches to representational analysis mostly modulates the level on which the computational core of cognitive science is further implemented. This implementational level of the cognitive research program is affected by embodiment in two ways. Traditional concepts are enhanced by new embodied concepts, and also resorted with respect to their importance in the new program.

I suggest that embodiment replaces Newell and Simon's metaphor of a general problem solver with Clark's new metaphor of the mind as **controller of embodied action** (Clark, 1997, p. 7). A shift of focus occurred in that the mind is not considered to be the "disembodied reasoning device" (Clark, 1997, p. 1) the problem solver used to be. However, the problem solver still has a place in embodied cognitive science, but just as one of the means that controllers of embodied actions have at their disposal.

This new picture of the mind as controller of embodied action takes in part means that we finally loose the restricted focus on symbolic representations that was already challenged by connectionism (e.g. Clark, 1989). First, the embodiment movement attempts to put symbolism and connectionism "in a proper perspective, so that the merits and demerits of explicit and implicit reasoning are each given their due" (Morley and Hunt, 2001, Ch. 15, p. 13). This point can be pushed even more generally: embodiment extends the traditional set of deductive strategies of symbolic reasoning over internal representations with inductive strategies of pattern formation over sensory motor-couplings. Second, the use of DST may introduce a third format for representations into the conceptual toolkit, in which representations can dynamically change over time as often found in the nervous system. Third, DST can give a law-like descriptions of some behaviour even when fully mechanistic explanations are not yet, or at all, possible. It can be useful to study how systems interact even if we do not know yet how they work on their own. At last, studying embodied actors does not look only at the internal structure of the system, but also includes the structure of their environments as object of study in the analysis.

Extending the toolkit of course opens serious discussions on the importance of the new strategies. Strong believers of embodiment might completely abandon disembodied high-level processes, whereas strong believers in traditional AI might regard embodiment as suitable explanation only for low-level processes. Clark takes an intermediate position. He explicitly rejects the second view, because he thinks it "is important not to conclude […] that facts about embodiment impact only our ideas about low-level sensory-motor processes. However, in the human case at least, we seem to find at all levels a mixture of highly 'embodied, embedded' strategies and apparently much more abstract and potentially decoupled strategies, with the creation and manipulation of external symbolic terms often functioning as a kind of bridge between the two" (Clark, 1999, p. 350). But still, he thinks that in embodied cognitive science "the tools and concepts to study complex, self-organising and embedded phenomena [are] prior to those of symbolic reasoning" (Clark, 1999, p. 350).

The practical level contains a diversity of changes

Most of the changes occur on a rather practical level as a consequence of implementing the computational mind as an action controller, rather than a problem solver.

Some of the changes are

- to study cognitive systems in the complexities of natural environment, rather than in toy environments of computer simulations, which allows
- to use the real world as an aid to problem solving, rather than insisting on the translation of physical quantities into symbolic objects, robbing the mind from possible environmental cues
- to use bottom-up approaches that use the theories of self-organisation, including complex response loops that couple real brains, bodies, and environment (Clark, 1997, p. 1)
- to focus on action loops between organisms and world which lead to more domain-specific and context-dependent knowledge rather than abstract and domain-independent knowledge, such as local and action-oriented inner representations that are (only) locally effective in guiding behaviour
- the use of epistemic actions, that use the world as "its own best model" (Brooks, 1991, in Clark, 1999, p. 350) within problem solving strategies
- to include the body as part of the computational loop
- the use of online rather than offline reasoning strategies, building robots rather than chess programs
- to emphasise how abstract reasoning might grow out of concrete reasoning (Morley and Hunt, 2001, Ch. 15, p. 13)
- the use of dynamical systems theory as a complementary tool to computational analysis

In conclusion, all these changes at the surface may render embodied cognitive science look widely different from traditional cognitive science. However, the research program continues the cognitive science tradition, as the embodied approach holds on to the computational core that defines cognitive science and aims to combine the new methods with the old.

4.3.2 Radical embodied research programs abandon computational analysis

In my understanding, the common core of radical views on embodiment abandons representational analysis and can be summarized by a few points. First, there seems to be a notion that the world is analogue rather than digital, and that the digital tools of computation and representation therefore in principle will fail to capture analogue processes adequately (e.g. van Gelder and Port, 1995, stressing on the continuous nature of time, Thelen and Smith, 1994, stressing continuous reciprocal causation). Part of this line of thought is that it is not enough if digital functions can approximate analogue processes with arbitrary precision. Roughly, because digital processes still in principle are of different nature than analogue processes, and that this difference is important.

Second, there is a quite radical appraisal that cognition is guided by action rather than by internal representation (Thompson and Varela, 2001; Varela, Thomson and Rosch, 1993, p.

172ff). Therefore, again, an analysis that starts with representation may be the wrong method to understand cognition. Some have taken this view as far as to discuss fundamental computational mechanisms such as memory (e.g. Clancey, 1997, p. 3; Glenberg, 1997) or category representation (Clancey, 1997; Lakoff and Johnson, 1980) as sensory-motor couplings rather than internal representations.

Third, intelligent action is understood as a non-separable result of continuous reciprocal interplays between complex systems and complex environments (Thelen and Smith, 1994; Varela et al. 1991). And again, either for principle or for mere practical reasons, they think that representational analysis likely will fail.

However, at current I consider radical embodiment to be more a theoretical position, rather than a (full-fledged) experimental program. In Lakatos's terminology, I think at present the radical program reaches only the status of 'immature science' (Winograd and Flores, 1987, p. 24) which is still in its "warm-up" period (John, 1998, p. 7) where it must develop an adequate conceptual language (including formal and mathematical tools) for empirical research. Until such a language is developed, empirical research will remain 'a mere patched up pattern of trial and error'" (Winograd and Flores, 1987, p. 24). There are several reasons for this judgement. First, these radical views seem to be guided by philosophical considerations on the nature of man and world, rather than by empirical necessity. Second, although theoretical considerations are a sensible starting point for any new research program, to me it is not yet clear how practical research in a radically embodied program will look like. This is partly because dynamical systems theory as the most prominent tool for experimental research in such radical programs is a) not worked out well enough to capture all of cognition, and b) only yields lawful descriptions of behaviour rather than full functional explanations. Therefore, its scope as solitary tool in cognitive science currently faces limitations. In sum, the conceptual language of the program needs further development at least in these respects[15].

Having said this, following Lakatos one should however not take an overly sceptical view in judging the radical program. Even if I would currently regard it as not having reached that stage of mature science yet, this still does not necessarily render the program less valuable than other research programs. First, the program establishes fundamentally different ways of thinking about the nature of intelligence. It thus attempts to ground research on the mind from the background of a fundamentally altered set of core assumptions. As I noted at the end of Section II, Lakatos's understanding of research programs leaves room for (motivated) *choice* especially at the level of the metaphysical core, since he thought that the core of a program could not be empirically justified.[16]

Second, new ways of thinking might seem to be of little value when viewed from the old perspective, especially if they can not show vast amounts of empirical results yet. This, however, might be due to the fact that creating a new core in part means to change some of the very background *needed to understand* a program. By definition, radically alteration renders (parts of) the new core incompatible with the old. If it is incompatible with it, the new core can thus possibly not be understood, less appropriately judged in its value out of the old

[15] However, Beer (2000) thinks that the program is on its way, because "dynamical approaches are beginning to engage substantive empirical questions in cognitive science" (Beer, 2000, p. 97).

[16] And to me it seems that the radical embodimentist have theoretical arguments that are worth consideration.

perspective. In line with this, Lakatos thinks that new research programs that show potential need protection in the beginning so that they can sufficiently develop to be judged in their value against competing programs. According to Bechtel,

> "one of Lakatos' important insights … is to recognize that in its early days a new program will not yet have achieved as much success as older programmes. Moreover, given the initial, simplified versions of the early theories offered in the programme, there will be many phenomena that seem to falsify them, but that subsequent theories in the program will nonetheless be able to handle if they are allowed to develop" (Bechtel, 1988, p. 62).

Interestingly, perhaps, Bechtel (ibid) thinks that both cognitivism and connectionism were granted with such warm-up periods in their early days. In sum, then, the demanding task of the future for the radical embodied research program will be to leave its conceptual warm-up period and migrate towards a grown up experimental research program. Otherwise it is likely poised to degenerate.

5 CONCLUSIONS

The purpose of this paper was to compare embodied cognitive science with traditional cognitive science. My main conclusion supports the view that embodied cognitive science splits in two versions. The weak version of embodied cognitive science shares the set of fundamental core assumptions with traditional cognitive science, but supplements them with more embodied and emergent explanations. The new embodied strategies include concepts to exploit the cues for action given by the environment, such as local and action oriented representations or epistemic actions which reduce the complexity of both representational content and processing. They include domain-dependent components organized in subsumption architectures, which are computationally cheap because they are adapted to specific problems. They also add dynamical representations that can change over time as a new format to symbolic and connectionist representations, and more generally use dynamical systems theory to describe the interaction of many components spread across the system and the environment at a larger scale. In sum, these tools are suited to study time critical behaviour online and in real worlds, in contrast to the traditional offline reasoning strategies that were applied to artificial environments. However, these concepts supplement and enhance the traditional concepts of computation rather than dispelling them. By contrast, radical embodiment completely abandons computation as an explanatory principle for intelligent action in favour of emergent, continuous reciprocal explanations of a world that is considered to be analogue.

My second conclusion concerns the status of weak embodied research. I consider it likely to be more powerful than the traditional AI program. First, the discussion has shown theoretical as well as empirical plausibility for embodied strategies. Second, the weak program is likely to gain in explanatory power over the traditional program because it aims to integrate embodied with disembodied mechanisms, giving both their virtues in the tasks they are best suited for. The weak embodied program can thus be expected to progress faster than the traditional program.

My third conclusion concerns the status of radical embodied research. I think it is at current rather a theoretical position in search for its own conceptual language needed as a basis for empirical research than a mature experimental research program. However, as it aims to assemble a research program out of fundamentally new ways of thinking about cognition, one should grant it with a ‘warm-up’ period needed to develop this conceptual level before systematic empirical research can start.

6 LITERATURE

Anderson, J.R. (2000) Cognitive Psychology and Its Implications, fifth edition, New York: Worth Publishers

Ballard, D. (1991) Animate vision. Artificial Intelligence, No 48, p. 57-86

Bechtel, W. (1988) Philosophy of Science. An Overview for Cognitive Science, Hillsdale, New Jersey: Lawrence Erlbaum Associates

Bechtel, W. (1998) Representations and Cognitive Explanations: Assessing the Dynamicist's Challenge in Cognitive Science, *Cognitive Science*, No 22, p. 295-318, http://www.artsci.wustl.edu/~bill/REPRESENT.html, as of September 2002

Beer, R.D. (2000) Dynamical approaches to cognitive science, *Trends in Cognitive Science*, Vol 4, No 3, p. 91-98.

Chalmers, D. (1992) Subsymbolic Computation and the Chinese Room, in Dismore, "The symbolicist and connectionist paradigms: closing the gap"

Clancey, W.J. (1997) Situated Cognition. On Human Knowledge and Computer Representations, Cambridge: Cambridge University Press.

Clark, A. (1989) Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing, Cambridge, Mass.: MIT Press

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

Clark, A. (1998) Embodied, situated, and distributed cognition, in Bechtel W. and Graham, G. (Eds) (1998) Companion to Cognitive Science, Oxford: Blackwell Publishers

Clark, A. (1999) An embodied cognitive science?, *Trends in Cognitive Science*, Vol. 3, No 9

Claxton, G. (2000) Hare Brain, Tortoise Mind. How intelligence increases when you think less, New York: The Ecco press.

Damasio, A.R. (1996) Descartes´ error: emotion, reasons and the human brain, London: Papermac

Dreyfus, H.L. & Dreyfus, S.E. (1986) Making a Mind versus Modelling a Brain, in Boden, M.A. (Ed.) (1990) The Philosophy of Artificial Intelligence, Oxford: Oxford University Press

Dreyfus, H. L. (1992) What Computers Still Can't Do: A Critique of Artificial Reasoning. Cambridge, Mass.: MIT Books.

Dyan, P. and Abbott, L.F. (2001) Theoretical Neuroscience. Computational Modeling of Neural Systems, Cambridge, Mass.: MIT Press

Eliasmith, C. (1997) Computation and Dynamical Models of the Mind. *Minds and Machines*. No. 7, p. 531-541, http://www.artscu.wustl.edu/~celiasmi/Papers/dynamics.mm.html, as of September 2002

Eysenck M.W. and Keane M. (2000) Cognitive psychology: a student's handbook, forth edition, Hove: Psychology Press

Fodor, J. (2000) Why the mind doesn't work that way. The scope and limits of computational psychology, Cambridge, Mass.: MIT Press.

Gardner, Howard (1985) The mind's new science. A History of the Cognitive Revolution, New York: Basic Books

Glenberg, A.M. (1997) What memory is for, *Behavioral and Brain Sciences*, Vol. 20, 1-55.

Hunt, G.M.K. (1999) Philosophy of enquiry, Department of Philosophy, University of Warwick.

Hurley, S. (1998) Consciousness in action, Cambridge, Mass.: Harvard University Press

Jaeger, H. (1998) Today's dynamical systems are too simple, in The dynamical hypothesis in cognitive science, *Behavioral and Brain Sciences*, No 21, p. 643-644.

John, R.S. (1998) Methodologische Probleme der Verhaltensbasierten Künstlichen Intelligenz unter kognitionswissenschaftlicher Perspektive, Magister Thesis in Computational Linguistics and Artificial Intelligence, University of Osnabrueck

Lakoff, G. and Johnson, M. (1980) Metaphors We Live By, University of Chicago Press

Markman A.B. and Dietrich E. (2000) Extending the classical view of representation, *Trends in Cognitive Science*, Vol 4, No 12

Morley I.E. and Hunt G.M.K (2001) The Philosophy of Psychology: the design and implementation of theories, Departments of Philosophy and Psychology, University of Warwick.

Newell A. and Simon H.A. (1972) Human Problem Solving, Englewood Cliffs, N.J.: Prentice-Hall.

Newell, A. and Simon, H. A. (1976). Computer Science as Empirical enquiry: Symbols and Search. *Communications of the ACM*, 19(3), pp. 113-126, March.

Port, R.F. and van Gelder, T. (Eds.) (1995) Mind as Motion. Explorations in the Dynamics of Cognition, Cambridge, Mass: MIT Press.

Russel S.J and Norvig P. (1995) Artificial intelligence: a modern approach, Upper Saddle River: Prentice Hall

Searle, J. (1980) Minds, Brains, and Programs, *Behavioral and Brain Sciences*, No. 3, p. 417-424.

Stillings et. al (1995) Cognitive Science. An introduction, Cambridge, Mass.: MIT Press.

Thelen, E., Smith, L. (1994) A Dynamic Systems approach to the Development of Cognition and Action, Cambridge, Mass.: MIT Press

Thelen, E., Schöner, G., Scheier, C. and Smith, L.B. (2001) The dynamics of embodiment: A field theory of infant preservative reaching, *Behavioral and Brain Sciences*, No. 24, p. 1-86.

Thomson, E. and Varela F.J. (2001) Radical Embodiment: neural dynamics and consciousness, *Trends in Cognitive Science*, Vol. 5, No 10, October 2001

Van Gelder, T. (1995) What might cognition be, if not computation? *Journal of Philosophy*, Vol 92, No. 7, p. 345-381

Van Gelder, T. (1998) The dynamical hypothesis in cognitive science, *Behavioural and Brain Sciences*, No. 21, p. 615-665

Van Gelder, T. and Port R.F. (1995) It's about time: An Overview of the Dynamical Approach to Cognition, in Port R.F. and van Gelder T. (eds.) (1995), Mind as Motion. Explorations in the Dynamics of Cognition, Cambridge, Mass.: MIT Press

Varela, F., Thompson E., and Rosch E. (1991) The embodied Mind: Cognitive Science and Human Experience, Cambridge, Mass.: MIT Press.

Winograd T. and Flores F. (1987) Understanding computers and cognition. A new foundation for design, Norwood, New Jersey: Ablex Publishing Corporation

Ziemke, Tom (2001) Are Robots Embodied? Invited paper at the First International Workshop on Epigenetic Robotics: Modelling Cognitive Development in Robotic Systems, Lund University Cognitive Studies, Vol. 85, http://www.ida.his.se/ida/~tom/papers.html, as of September 2002