Holm, P., & Karlgren, K. (1996). Cognitive Science on Trial. Proceedings of the 19th Information systems Research seminar In Scandinavia (IRIS), Lökeberg, Sweden.

Cognitive Science On Trial

Peter Holm and Klas Karlgren
Department of Computer and Systems Science, Stockholm University,
Electrum 230, 164 40 KISTA, Sweden; Tel: +46 8 16 1684,
E-mail: {pholm, klas}@dsv.su.se

Abstract:

For more than twenty years, theories about human cognition have been applied in various areas of computer science. During the last years these theories have lost much of their original currency, as substantial criticism has been formulated against traditional cognitive science. At the same time, this criticism has resulted in new approaches, sometimes collectively associated with the term "situated cognition". We believe there is a need for a deeper reflection on how cognitive theories can be applied within information systems research.

In this paper we present a framework for discussing the various criticisms of traditional cognitive science. We emphasize that there exist fundamental problems regarding the very idea of having a science of the mental. As a consequence, we problematize also the new approaches, and criticize the criticism. Instead we argue for a practical and eclectic evaluation of all theories, including the traditional one. At the core of the argument is an attempt to free the field from its philosophical ambitions. The whole discussion must center around practical issues related to the design and usage of information technology. At the end of the paper, we discuss concretely what cognitive theories could be applied in what practical situation.

1 Background

In the fifties, when the first computers were built, cognitive science emerged with an ambition to create artificial intelligence. Intelligence was here seen as the physical manipulation of (hardware-independent) symbolic representations. Ever since these days, the computer has served as a dominating metaphor to describe human intelligence.

When cognitive theories have been used within computer science, much work has been motivated by a hope to capture parts of human thinking in formal models. In the eighties, considerable investments were made to construct expert systems to replace or support various types of human problem solving activities. Cognitive theories have also been applied in interface design, e.g., when studying human abilities to understand information presented in various ways. As such they have played a general role of capturing the human part of human-computer interaction. Much criticism of cognitive science, is focused on the appeal to the traditional theoretical approach. Hence we center our discussion around this theme.

The rise of modern cognitive science can be associated with names like Newell, Simon, Chomsky and Fodor. They all viewed cognition as a manipulation of symbolic representation of the world. We will refer to this view as "the representation hypothesis". In the end of the eighties these ideas were seriously questioned in the fields of cognitive science and computer science (Winograd and Flores, 1986; Suchman, 1987).

Going through the history of Western philosophy we would find a wide variety of versions of this basic theme. For the purpose of this discussion, we would like to distinguish between the following variants: (a) Human cognition is described in terms of *representations describing a portion of the world*, e.g., in terms of propositions, ideas, or mental images. (b) Human cognition is described in terms of *symbol manipulation* (also non-descriptive ones), e.g., in terms of physical control systems, deductive systems, or signalling. All arguments against the hypothesis that we will discuss below apply to at least one of these two alternative formulations, but not always to both.

2 A Framework for Understanding Its Failures

Below is a framework for discussing the various criticisms against the hypothesis.

A Criticism prompting for extentions/alternative theories

- i The hypothesis is based on a *mechanistic view* on human action and fails to consider its situated character.
- ii It *over-rationalizes* human behavior and assumes that human cognition can be described in isolation from, e.g., emotional and bodily aspects.
- iii It is based on an *individualistic perspective* and disregards social aspects.

B Problems with the very idea of mental representations

C Problems with the very idea of having a science of the mental

In sections three to seven we will comment on each of these themes (A.i, A.ii, A.iii, B and C) respectively.

3 Mechanistic Versus Situated Actions

The first theme concerns the mechanistic perspective on human activities underlying the representation hypothesis. As an alternative, Suchman (1987) has suggested an approach associated with the term "situated actions".

A possible misunderstanding of this criticism, is that it prompts for a simple extension of the traditional approach to capture more situational factors. This, in turn, may be achieved through the introduction of more complex mental representations that include more factors of the situation. Such a criticism of the

representation hypothesis will of course be fallacious, since the result will only be that the theory is made more fine-grained and broader in scope. It does not change the basic character of the theory.

Disregarding crude misinterpretations like these, we believe that the term "situated" is open for a set of different interpretations also in its original use. This is perhaps one of the reasons it has become so widespread. It is related to (at least) two important observations about human activities: Human actions are always situated in (a) a contingent physical context, and (b) a social and cultural context.

These observations can have a set of different implications for the study of human cognition, depending on how we expand the argument. One of the things Suchman (1987) sets out to study, is how different participants in a dialogue have different access to their immediate physical environment, and how this affects their understanding of each other. She applies these ideas when she analyses the dialogue between a user and a Xerox machine. She tries to reveal assumptions made by either party on what has happened with the papers and the machine. She emphasises that (the designer of) the machine has a limited knowledge of these things, and that this is an obstacle to overcome in the design of the machine.

Another way in which the argument can be expanded, is to attack the *mechanistic view on human actions* underlying the traditional approach. Suchman (1993) summarizes the core of her critique as an attack on the assumption "that the relation of plans to actions [is] assumed to be... a determining or controlling one". (This argument is of course applicable also to theories about other types of mental phenomena.) As an alternative to this view, she suggests a perspective where human actions are viewed as "constant improvisations". Background knowledge, plans, social norms, et cetera, may serve as a resource for these actions, but not as a controling system. Instead of explaining human actions mechanistically she retains a perspective on each human as a sovereign and responsible actor.

4 Intellectual Contemplation Versus Emotions and Bodily Instincts

Another line of criticism against the representation hypothesis is here labeled "over-rationalization" or "over-intellectualization". Traditionally human activities are explained in terms of abstract symbols. When this is applied to ordinary everyday activities, it leads to an over-emphasis on the role played by intellectual reflections in our lives. At the same time it de-emphasizes bodily and emotional aspects.

To explain human activities in cognitive terms, can be associated with ascribing psychological or mental states and processes to an individual. This

ascription can be associated with different claims as to the psychological or mental reality of that individual. In psychological think-aloud studies, a report that a subject contemplates a graphical diagram is typically interpreted as indicating a conscious reflection upon such a diagram. Such a study may be aimed at finding generalizations about different ways of thinking. In other situations, references to mental phenomena can be of a more metaphorical nature. We may, for instance, say that an infant knows that his/her parents exist. But what do we mean by that? Does the child reflect upon existing and non-existing phenomena and then categorize the parents into the latter category? Is this something that any infant ever consciously thinks about? Similarly, does the child "know" that the chair will hold his/her weight and not collapse? (This example is from (Wittgenstein, 1969).) Sometimes we even attribute mental states to animals. We may say about a dog, e.g., that "Old Fido knows it is time for a walk now". But this does not mean that Fido mumbles to himself "Here we go again, another one of these walks." (Malcolm, 1973).

Does the cat "know" that milk exists, or does it perhaps merely "have this belief"? (Another example from (Wittgenstein, 1969).)

Such "metaphorical" references to mental phenomena can be a way to talk about human activities from an observer's point of view. It can also be a way for the actor to reflect on his/hers actions. In either case, the vocabulary is not meant to reveal a psychological (or neurological) reality. This is how Suchman concludes her discussion on "background assumptions":

> A background assumption, in other words, is generated by the activity of accounting for an action when the premise of the action is called into question. But there is no particular reason to believe that the assumption actually characterizes the actor's mental state prior to the act. In this respect, the "taken for granted" denotes not a mental state but something outside of our heads that, precisely because it is non-problematically there, we do not need to think about. (Suchman, 1987, p. 47)

A version of the same theme is found in the discussion on skills (Anderson, 1980; Dreyfus and Dreyfus, 1986). It is pointed out by many, that the need (and often the ability) to give declarative descriptions of one's doings gradually diminishes when a skill has been acquired. Verbal rules may have a role to play in a learning phase or as a reflection on an already acquired skill, but what happens with the rule once the skill is acquired and the ability to account for the rule is lost? Does one have to assume that a (subconscious) representation has been "stored"?

Another example is the skilled expert, e.g., a professional tennis player. We do not need to postulate a "hard coding" of the verbal rules of an instruction on how to hit the ball with the racket. The skill can rather be seen as a pure bodily ability. Or as Frederic C. Bartlett put it in 1932:

> We may fancy that we are repeating a series of movements learned a long time before from a text book or from a teacher. But motion study shows that in fact we build up the

> stroke afresh on a basis of the immediately preceding balance of postures and the momentary needs of the game. Every time we make it, it has its own characteristics.
>
> [T]here is no reason in the world for regarding these [traces/schemata] as made complete at one moment, stored up somewhere, and then re-excited at some much later moment. ((Bartlett 1932) quoted in (Clancey, 1991, page 211))

We may also want to view certain actions as primitive and instinctive. When a child for the first time sees a dog barking and running towards it, and the child shields itself with the arms, is this behavior reasonably viewed as a result of an abstract reasoning process, where the dog's behavior is viewed in analogy with the child's experiences of other running big animals? Is not the child's behavior better viewed as an instinctive, unreflective action? (Malcolm, 1982)

There are, hence, good reasons for questioning the assumption that representations of some sort are involved in our everyday unreflective activities. The fact that we normally make metaphorical ascriptions of mental states or processes in such situations does not suffice as a justification.

Notice, though, that even if we made a distinction between reflective and non-reflective activities and refrained from talking about representations in the latter case, the problem still remains on what grounds we can talk about representations in the first case, when we actually do talk about conscious intellectual reflections. (This will be further discussed below.)

5 Individual Versus Social Aspects

Another theme in the criticism concerns the individualistic perspective of modern cognitive science, which leads to a disregard of social aspects. A problem with this criticism is that people can mean a set of different things when they talk about "social aspects". In its most general form, the discussion can be held on an ontological level. To be a human is to be a social being. As an example we may consider Heidegger's notion of Sorge. According to Heidegger, it is constitutive of the human being to be directed towards (to care about, to be related to) other humans. In this sense, no individualistic theory can ever hope to fully account for human cognition or any other aspects of the human being.

We suggest that the concern for theorizing about social aspects within cognitive science can be classified into four potential areas of interest: the collaborative construction of concepts, negotiation of how to use concepts in specific situations, the use of socially design artefacts to support thinking, and collective competence.

5.1 Collaborative Construction of Concepts

Why is it that skiers and Eskimos have so many words for snow, while ordinary people in Sweden have only a few? It seems to be related to the fact that these

people are involved in a set of social activities, where they need to talk to each other about different types of snow, e.g., when they learn how to wax and prepare their skis. Traditionally cognitive science has focused on describing the thinking, skills and knowledge of individuals in terms of representations and symbol manipulation. What is not covered in such an analysis is the social status of the way individuals are using particular concepts (or representations). With a certain set of concepts comes a way of thinking, a way of talking and a way of acting. But this is by no means something given for all times. New ways of thinking, talking and acting emerge all the time. Since language and cognition is so rooted in human activities it seems reasonable to assume that if we shall describe how people think, talk and act, we must also include a description of how their concepts and language are formed in a social context. Some researchers have taken this argument even further and claim that traditional mental concepts, such as "knowledge" and "learning" are more naturally ascribed to groups, societies or organizations.

Analysing social constructions of concepts, may involve studying social controversies and conflicts. When a person categorizes another person as dumb, creative, psychotic, smart, or as belonging to a certain ethnical group, this categorization is far from being something neutral. We may need to discuss what it means to use such concepts in specific situations, whether they are good or bad, what objectives are met with these concepts, why they are used, et cetera. From this perspective, it is far from illuminating to simply say that a person X "knows" that another person Y is neurotic, or that X has a mental representation of Y being neurotic. This would be to abstract away all crucial difficulties in the analysis of this concept. What complicates things further is that language and concepts are both the media and the object of our study. We must hence also critically question on what grounds researchers use a certain vocabulary.

The above type of objection to traditional cognitive science is not limited to a particular set of concepts that are especially problematic. The argument is in principle applicable to all concepts, e.g., to "table", "multiplication", "c-minor", "mountain", "football", "snow", et cetera.

5.2 Negotiation of How to Use Concepts in Specific Situations

Another way to study social aspects of cognition is proposed by Suchman (1987), when she sets out to investigate how the interpretation and conceptualization of a specific situation can be understood as "ensemble work", i.e., as a negotiation between two or more persons on how to use certain concepts in a particular situation.

The social construction of concepts is not something that is suddenly brought to an end. Rather, we are constantly involved in a re-shaping and re-interpretation of our shared social praxis. Every time we use words, we confirm

or re-shape old social conventions on how to use these words. Each new situation is unique and requires a unique interpretation. Most often this is handled intuitively, immediately and effortlessly. Sometimes, however, it may be problematic and require special attention.

Consider as an example a mother that asks her son whether he has cleaned his room. He answers yes. When she later enters his room she complaints that he has not done any vacuum cleaning. He on the other hand replies that he made his bed. In this situation they need to negotiate what counts as "cleaning the room". It would not be particularly illuminating to describe the son as having a mental representation of the fact that he has cleaned his room. Rather we must pay attention to the social problems of negotiating what these concepts mean in certain situations. This, in turn, is related to problems with power, dominance, social responsibility and changing roles.

5.3 Use of Socially Designed Artefacts to Support Thinking

A third way to talk about social aspects of cognition, is to study how human thinking is supported by socially constructed artefacts, e.g., rulers, calculators and computers. This focus is today prevalent in some new branches of cognitive science. It is associated with the term *distributed cognition*. External material artefacts like pens, paper and computer interfaces are, on this view, inseparable parts of human thinking. This discussion could be extended to include all kinds of social and cultural constructs, such as cities, work organizations, language, mathematics, et cetera (Dahlbom, 1987). Human cognition and human mind can then be seen as an "artificial phenomenon", in the sense that it consists of a mastery of social artefacts like these. With this terminology, Dahlbom turns the question about artificial intelligence upside down, by concluding that the human mind is artificial (Dahlbom, 1993).

5.4 Collective Competence

A fourth way to talk about social aspects of cognition, is simply that we sometimes need to view a group as such as having a certain capacity to think and to act. Arguably, the competence of the group can be something more than the "sum" of the competencies of its members. It may not be possible or desirable to describe exactly what is the contribution of each individual in the group.

Notice, that the first three ways to talk about social aspects concerned what social aspects are relevant to consider when describing individuals. In this fourth case the individuals disappear and we focus only on the group.

6 Problems With the Very Idea of Mental Representations

Some of the issues raised so far, have been discussed within philosophy for centuries. To summarize this discussion would be to summarize the core of over 2000 years of Western philosophy, since the philosophy of mind has had such a central role in this tradition. However, in this section we would like to mention a few philosophical problems with the the very idea of mental representations: (1) the infinite regress argument, (2) the template matching problem, (3) the private language argument, and (4) objectification of the mental.

(1) The infinite regress argument was formulated in the eighteenth century, when the British empiricists, e.g., Hume, Locke and Berkeley, formulated their ideas of "ideas". Essentially they believed that our ability to recognize that a particular green thing is green, depends on our ability to form mental "idea" of the color green. Once we have seen a set of green things, we acquire the idea "green", which is pretty much like a mental image of green. These philosophers soon realized that there were severe problems with this kind of explanation. One of the problems can be formulated as follows: If recognizing green depends on an ability to compare an impression of green with a (mental) idea "green", how does one recognize the idea "green" inside our heads? If recognition in general depends on an ability to compare things with ideas, then we need a second idea to recognize the first one. In this way an infinite regress appears. The problem has not been solved but only moved into our heads. Moreover, one may ask: Who or what inside our heads perceives the inner ideas?

A similar argument is often associated with Wittgenstein in his discussion on rule following. If our ability to follow rules is explained in terms of mental rules, then we will need another rule to tell us how this mental rule should be followed, et cetera. This argument can be generalized to any and all kinds of mental representations.

(2) The template matching problem is another argument against mental ideas and images associated with the British empiricists. It is based on the observation that a particular word or concept can be associated with zero or many different images and that each image can be associated with zero or many different words. Imagine for instance the word "dog" and various pictures of dogs, or a picture of a dog and the words "dog", "pet", and "warning". An image cannot, therefore, determine the meaning of a word, and, moreover, it cannot indicate how a word is used.

(3) The private language argument was formulated by Wittgenstein (1958). In short it argues that we can not refer to any private items, like mental representations, when explaining social phenomena of language. The criteria for correct language behavior must be determined socially. Wittgenstein further

argues that the very idea of a private language is contradictory. For a language to be useful, there must exist criteria for right and wrong language use. If we really imagine a private language, there could be no such criteria. No one can enter the scene and say: "You use that word (or representation of some sort) incorrectly.". In each particular situation the subject simply decides to use a certain word about a certain phenomenon. There can be nothing more to it than that. Strictly speaking, this is not a language, according to Wittgenstein. Why would we need such mental symbols? What would their functional role be?

Notice that this argument is not applicable to the idea that people utilize already existing socially formed concepts when they think.

Notice also, that in this discussion "social" is synonymous to "public" and contrasted with "private". A related problem concerns how we can talk about private phenomena at all. Is not the word "private" used to signify that this is something unavailable for others, e.g., my own unique experiences and thoughts? Wittgenstein solved this by concluding that all mental phenomena are talked about from a public point of view. Even if we have respect for each individual's unique experience, our talk about these experiences is based on socially available facts about them, e.g., how the individual behaves and what he/she says. Whatever a person's experiences are, our classification of that person as being courageous or cowardly is related to the external behavior of the person. The same holds for expressions like "being in pain", "solving problems", "understanding", et cetera.

(4) The objectification of the mental is the fourth issue we would like to mention. Many mental words seem to *refer* to some*thing*. Wittgenstein emphasizes that this must be viewed as a metaphor and that our mental vocabulary should not be regarded as descriptive in a literal sense (Malcolm, 1986). However, we often misinterpret our talk about the mental and the mind, so that we believe it is really referring to some thing inside ourselves. Hence we expect to find some mental states or some neurological patterns that correspond to these words. This is simply a misinterpretation of our language use, according to him.

Closely related to this, is the issue of *essentialism.* Concepts and entities are assumed to have an essence, i.e., that there are certain necessary properties which have to be present for an entity to exist, or to be what it is. For instance, whenever anybody remembers anything, it is often assumed that a certain process of remembering must have occurred. Wittgenstein pointed out that such invalid assumptions often underlie questions such as "What (kind of process) is remembering/learning/understanding et cetera?". Such psychological concepts may instead refer to families of phenomena and behavior, the members of which do not all have to share any common properties.

7 Problems With the Very Idea of Having a Science of the Mental

Now we have reached the last theme in our framework: the problems with the very idea of having a science of the mental. This theme can be turned into a criticism against the other criticisms, if these are based on the assumption that one can overcome the problems with the representation hypothesis by simply constructing new and more subtle theories of cognition, covering more aspects. We will present two arguments against the possibility to have an ultimate theory of human cognition. The first is based on Wittgenstein's conclusion that it is impossible to rationally justify our belief in other minds (and various other things we take for granted in our every daylife). We apply this from a third person perspective and conclude that it is impossible for a researcher to question and explain all aspects of another person's behavior and thinking, e.g., the other person's belief in the existence of other minds. The second argument concludes that there is a basic difference in how we relate to the material world and how we relate to other humans. In this difference lies that we do not manipulate and control other humans. Nor do we explain them in the same causal and deterministic way in the same way as we explain the physical world. If we explain humans, we will relate to these explanations differently compared to our explanations of the material world.

7.1 Digression on Wittgenstein's Foundationalism

The term foundationalism is here chosen to denote Wittgenstein's position, which is closely related to the "over-rationalization criticism", mentioned above. From this position follows that there are principal intellectual limitations in the philosophy and research about human mind. This is related to the fact that science itself is based on a primitive acceptance of certain ways of living, which can not be questioned or studied within science. In what follows we will largely depend on (Wittgenstein, 1969) and (Svensson, 1991).

The core idea here is that the foundation of rationality is not itself something rational. It is something we accept without argument or intellectual justification. It is visible in our everyday behavior, e.g., when we make judgments like "This is a hand" or when we say "My name is Kurt". We do this intuitively and immediately, without reflection. If someone fails to make such statements, we will have problems communicating with that person, since parts of what we normally take for granted would be lacking.

According to Wittgenstein, many classical philosophical problems arise because philosophers set out to to give this foundation a rational motivation, e.g., when they set out to prove the existence of the physical world or the existence of other human minds. According to him, it is impossible to doubt, to prove, to give

evidence for, or to know these things, unless we seriously change the way these words are used (i.e., the words “doubt”, “prove”, “know”, et cetera). Our trust in the existence of other people is a prerequisite for doubting, proving and knowing. If we should try to justify that, what could count as evidence? What is more reliable? Rather, our acceptance of such “facts” is part of the foundation of knowledge and truth. Hence, strictly speaking, these statements are not true or known themselves. (Strictly speaking they are not facts either, but we must call them something.) Wittgenstein called such indubitable facts “objectively certain”.

This leads to the shocking conclusion that science, philosophy and all other intellectual activities are hopelessly embedded in real life(!). They are based on an intuitive, immediate and arational way of living, where people simply take a set of things for granted without even bothering about giving them a rational justification. When a child is born, it does not intellectually motivate its trust in the mother’s existence. It simply does this without doubt. Doubting and knowing comes much later in life. The primitive acceptance comes first and underlies these other more advanced activities. The intellectual philosopher can not shake this primitive background off her back. She is doomed to be both a primitive animal and an intellectual being. Indeed we may conclude that it is a fallacy of the philosophers to see a contradiction between the two.

For Wittgenstein these reflections led to the conclusion that it is impossible to construct metaphysical theories about the foundation of knowledge and language. In this respect his philosophy differs from, e.g., Aristotle’s theory about natural categories, Kant’s theory about our forms of perceiving the world, and phenomenological theories about the human mind and its object (intentio and intentium).

Let us return to the basic theme of this paper. How is foundationalism related to the representation hypothesis? We believe it is relevant in two respects: *Firstly*, it can be used to strengthen the argument against over-rationalization, mentioned above. How rational are people? On what grounds can we assume that abstract symbolic representations underlie our ability to act and think? *Second*, and more importantly, we believe it can be used to make general reflections on cognitive science (and philosophy) as such. How rational are the researchers? Arguably, they share the same foundation as their objects of investigation. They can not distance themselves from it and study it from a neutral position. They can not, as Descartes attempted to, throw themselves as lonely subjects out in an empty space, seeking a purely intellectual and rational foundation for their analyses. How a person comes to believe in physical objects, is a question unavailable for them. What alternative beliefs could there be? The primitive acceptance of this and other facts (and a blindness for them) is a prerequisite not only for the humans under observation but also for the researchers. Before

becoming researchers, they must accept this. In an important sense, this foundation can not be explained or studied within science, and our scientific explanations of human cognition can never be made complete. Even if we can produce some rational models of human behavior in terms of symbolic representations, these models must be based on some concepts for which there exists no justification, concepts that are understood intuitively and arationally by the researchers, as a part of life.

7.2 Science, humans and nature

We concluded above that science is embedded in our ordinary lives. It is only one out of several activities in which we are involved. Our intellectual activities have no privileged position in the sense that all other activities are based on these, somehow. Rather it is the other way around. Once we have realized this, we may go on and argue that there is a difference and a potential conflict between science and other activities in our lives, e.g., our way of relating to other humans.

The paradigm case of science is to give deterministic and causal explanations of various material phenomena. Once this is done, we may predict, control and manipulate these phenomena. This, however, is not the way we treat other humans. Sometimes we manipulate them, but if this reaches a certain level, we normally feel it is unethical. Rather our relationship towards other humans is characterized by having respect. It lies in our conception of a human subject that a subject has a free will and is unpredictable. If we accept this view, it follows that we will never succeed in giving deterministic and causal explanations of human behavior, in the same way as this is done in natural science. At least it can not be done without seriously changing our conception of causal explanations or our conception of human subjects. Nor will we succeed in reducing our talk about mental phenomena to our talk about the material and physical reality. Arguably we have totally different ways of behaving towards the material world and towards other humans.

Notice that the point here is not that we choose to treat humans in a certain way and that we choose to treat the material world differently. This is not something intellectual. It is something we simply do. We have no other alternative. The child does not choose to treat its mother as a human, and not as a machine. How could it do otherwise? This is a behavior that we do not question or justify, and in a certain sense we *can not* give it a justification. It is simply another part of our primitive arational foundation of rationality, besides our acceptance that there exist physical objects.

The problem with mechanistic versus non-mechanistic perspectives on cognition is hence a version of a general problem for the social sciences and the humanities. Moreover, this problem will not be solved by simply extending the

traditional approach in cognitive science to include also bodily and social aspects. Also these theories can have a mechanistic perspective on the individual subject.

Unfortunately, we believe that cognitive science from the very start has been burdened with a desire to solve traditional philosophical problems. We also believe there is a risk that the newest research directions, focusing on social and bodily aspects, inherit these ambitions, trying to play the same role but with a new and "better" theory. If we accept the argumentation above, however, it follows that there exist philosophical "problems" with the mental that *can not* to be "solved". Human subjects can not be explained. The language about mental phenomena can not be reduced to a language about the physical (and social) reality. The only thing we can do is to give up such ambitions. What we can hope for is that we can do this with peace in our minds, with a clear understanding of why the problems can not be solved.

What can we do in cognitive science then? Much. Once we have given up the intellectual hubris that underlies this discipline from the start, we can start to appreciate the contributions within the field from a practical and less pretentious point of view. However, if we really take the above argumentation seriously, there is no fundamental difference between the new research directions and the old blunt and "naive" model of mind as a manipulation of symbolic representations. None of these theories are "complete". They all focus on a limited set of phenomena. What particular aspects should be covered in a specific situation, can only be justified by reference to concrete practical problems. We can no longer appeal to an ultimate "truth" about human cognition. We can not claim that theories of cognition must cover social aspects, simply because it is possible to do that or because it is shown from some perspective or another that social aspects affect the way we think. There will always exist such aspects "just around the corner" for every theory of cognition. It will always be possible to extend current theories to cover more phenomena. Such an extension can be justified practically, but we can never hope to solve the philosophical mysteries of mind in this way.

8 Practical Implications for Research within Cognitive Science

Let us now return to the various research directions within cognitive science and explore their contributions on a more practical level. Based on the framework presented in the previous sections we can distinguish between the following potential areas of interest: symbolic representations, situational aspects, e.g. real-time interpretation of the physical contingent environment, bodily aspects, social aspects, e.g., collaborative construction of concepts, collaborative application

negotiation of how to use concepts in specific situations of concepts, use of socially designed artefacts to support thinking, collective competence. In what follows we will exemplify how these areas can be studied practically.

8.1 A Revival of the Representation Hypothesis

Consider as an example some researchers studying how fast rats can learn to find their way out of a maze by giving them various types of re-enforcements. In this example, the old traditional representation hypothesis is well suited. Whatever experiences rats may have, the only thing the researchers are interested in is their external behavior. A convenient way to do this is to conceptualize the rats as navigating by means of mental maps of the maze. Similarly, errors are described in terms of fallacious mental maps. The representation hypothesis is here simply a metaphorical way to talk about the rats' ability to orient themselves in the maze.

Another example is using IQ-tests to measure people's ability to solve spatial problems or to make mathematical calculations in their heads. The subject's results can be measured against social criteria for correct manipulation of symbols. Hence it is reasonable to view the subjects as performing *something that corresponds* to these external methods for symbol manipulation. How shall we view this ability? How shall we describe it? The simplest alternative is of course to talk about it as a mental or cognitive version of the external manipulation of symbols. This way of talking can not be taken literally, but it is a forceful and economic model, well suited for the purpose at hand.

We can also use the perspective of distributed cognition to reinterpret the usage of the representation hypothesis. Intellectual and cognitive abilities are nothing but abilities to manage social artefacts, like mathematics and language. There is nothing more to it, and we normally talk about these abilities in terms of these artefacts.

In this example it is not necessary to bring in social aspects of cognition. We are not, e.g., interested in describing the cultural history of Euclid's algorithm. Rather our research objectives concern individual variations that can be measured, *given* that all involved parties accept modern mathematics as the norm. The social status of Euclid's algorithm can hence be made invisible and held constant. The algorithm has been used for over 2000 years. Most individuals in modern societies show a strong consensus on how to use it. If we can not take this for granted in scientific studies, what can we take for granted?

8.2 Collaborative Real Time Interpretation of Environment

Consider as another example a medical expert team giving telephone support to ambulance staff. How could one evaluate their work praxis or write guidelines

for their job? Arguably there will always exist judgments regarding the patients that need to be discussed in each unique situation. It is not possible to formulate rules on beforehand prescribing how to make all these judgments or how to give instructions for the judgments. In this example, it is important to have respect for the uniqueness of each situation and the need for a collaborative discussion on how it should be interpreted.

Related to this collaboration, we may need to pay attention to roles, responsibilities and skill requirements for the various work groups involved. (Of course, every situation is contingent and needs a real time interpretation, but people show different degrees of consensus regarding different types of judgments. That is what makes this example different from the example with Euclid's algorithm.)

In this example it is not, however, necessarily important to discuss the social and cultural construction of the concepts used. Hence we need not study how modern medical terminology is formed within various research communities involving miscellaneous types of social controversies.

8.3 Skills and Embodied Semantics

Learning and reflecting on skills, is an example where it may be necessary to focus on bodily aspects. You can not learn to play tennis or an instrument from only reading books, without practising bodily skills. Moreover, it may require certain bodily skills to understand certain terminology, e.g., when a piano player says that a melody should be played with "whipped cream in the hands". The study of this type of dependency between semantics and bodily abilities is often labeled "embodied semantics" and can be applied also to more common phenomena like the word "chair" and our ability to sit.

A focus on social praxis may be needed when there are no available social norms for symbol manipulation, but only norms for evaluating practical skills. In learning situations, one also need to focus on praxis as a complement to abstract symbolic thinking, e.g., by giving concrete examples in mathematics. Consider as an example, students learning to solve differential equations. Here we can analyze the way they should think in terms of mathematical algorithms. There exist social criteria for right and wrong manipulation of symbols. It may not, however, be a trivial task to apply these learned algorithms to concrete practical problems. The use of these algorithms in practical physics problems involving speed an acceleration requires additional skills and training.

In other situations we may have no language at all to talk about people's thinking, even though there exist clear social criteria for correct behavior related to the practical activities in which they are involved. Consider as an example a nurse judging whether a patient needs special attention. There may exist criteria that can be applied afterwards for checking whether it was necessary, e.g., by

making certain tests. At the same time there may not exist concepts (explict rules) describing how nurses should make such a judgment. The judgment can not, with existing terminology, be described in terms of symbol manipulation. There may exist certain concepts, such as “colour of skin”, “degree of consciousness”, et cetera, that can be used to give certain instructions. It may not, however, be possible to define an algorithm for the judgments based on these concepts. Moreover, the use of these concepts may also require certain skills and training.

8.4 Connectionism

Consider yet another example, where we want to develop a computer system that predicts faults in a telephone exchange, given a set of measurement data. There are experts in the field at our disposal. We know what data they take into account, but we can not describe their predictions in terms of rules over these data. In this situation we may use a connectionist model of their expertise. Treating their guesses as normative, we program a “neural net” with examples. (We could also give real solutions as examples to the program, but then it would not be used as a model of the expert’s skill.)

In some branches of contemporary cognitive science, connectionism seems to have inherited much of the (abnormal) expectations that were directed towards artificial intelligence (AI) in the eighties. We feel it is hard to understand why. The approach has a set of benefits over traditional AI, primarily due to its negative features, i.e., what it is not. It is not based on explicitly stated rules. However, it also shares a set of limitations with the traditional representation hypothesis. Primarily is *mechanistic* and *individualistic*. Some claim it covers *bodily aspects*, since it concerns neurons. On the other hand it may also be criticized for ignoring bodily aspects, i.e., all aspects but the neurons. Every bodily phenomenon that does not result in a neuron firing is viewed as irrelevant for thinking, e.g., heart beating, breathing, hormones. At a closer look, the connectionist model of mind shows many similarities with the representation hypothesis. The major difference is perhaps that the former focuses on another level of digitization — the neurons. Moreover, as the model is used in this (and many other) examples the input and output consists of high-level cognitive symbols. It is hard to see how these are meant to match the neural level.

8.5 Collaborative Construction of Concepts

Let us now consider a development of an expert system for portfolio management at a bank. During a phase of knowledge acquisition the developers analyze a set of experts. A new categorization of companies and business sectors is introduced, together with a standard method for minimizing risk and

optimizing profit. The idea is that advisors should use this standard method as a starting point in their discussions with customers, adding comments and judgments based on their own expertise. In this situation, one can not ignore the social aspects. New concepts are introduced. The company is involved in a refinement of their expert terminology. This terminology does not yet have a stable usage within a social praxis. The new analysis method will probably lead to new work roles, changed responsibilities and changed skill requirements for people working with portfolio management. All this needs to be discussed and evaluated. Needless to say, it would be fatal to use the traditional individualistic representation hypothesis as a theoretical foundation here. Such an approach would lead to a narrow focus on experts in isolation from their organizational context, describing their expertise in terms of a manipulation of symbols, disregarding the need to established a shared social understanding of these symbols.

9 Practical Implications for Research within Computer Science

Some of the conclusions above concerning cognitive science, can be directly transferred to a discussion on the status of cognitive theories within computer science. Typically cognitive theories are used in the design of help systems, explanations, tutoring systems, interface design, et cetera. We believe that a narrow individualistic focus on symbol manipulation is reasonable to address many of these issues. The problem is to know when. We need to identify criteria for choosing a certain type of theoretical foundation in a particular situation. One such criterion is to ask whether the research problems really concern *abilities of individuals* to understand and use certain systems, *given that the social use and basic functionality of the system are accepted*. The aim may, e.g., be to find general design guidelines as to how information and system functionality should be displayed to the users. Navigation in hypertext systems is a good example. We believe that *the traditional approach* in cognitive science will continue to play an important role to address such issues, since the notion of mental representations is a convenient way to talk about people's cognitive abilities here.

Also the example above about *real time interpretation of environment* has its correspondence within computer science. These factors are important, e.g., in the design of systems used in critical *risk situations*, like monitoring systems in nuclear plants. Important research questions then concern how you can construct a dialogue with the user to "negotiate" how the current situation needs to be interpreted, what additional checkings need to be done, et cetera.

The concern for *skills and praxis* are also important, e.g., when finding a *balance between abstract* explanations and instructions through computer use,

and *practical involvement* in particular situations. Here computer based *simulation* and *multimedia* interfaces can play the role of bridging this gap. Another way to bridge this gap is to give concrete examples and illustrations.

A crucial success factor for *connectionist approaches* (and machine learning in general) concerns the identification of *normative examples*. The individualistic perspective underlying this approach, may blind us for important social problems here. Imagine a neural net used to capture the skills of a person judging loan applications in a bank or a person analysing the stock market. To use such a system is to put trust in these persons. But, on what grounds do they make their analyses? What alternative perspectives are there? Are there hidden social conflicts underlying their judgments? Moreover, there is a risk that the lack of explicit rules makes these concerns even more invisible to us.

One way to find situations with clear uncontroversial normative examples, is to focus on games and artificial block worlds, avoiding the complexity and controversy of real life. But, is the utility of connectionism really limited to such situations? We feel it is important to critically problematize the use of connectionist approaches (and machine learning) in "real life applications" from a social point of view.

Another area where we believe an individualistic perspective has caused severe problems, is the field of *AI* and *expert systems*. Much terminology and many methods are based on the representation hypothesis. Individual experts are analyzed during a phase of "knowledge acquisition". Knowledge is seen here as something possessed by individuals, describable in terms of symbol manipulations. In this situation, we can use the above framework as a guideline for finding potential breakdowns of the representation hypothesis. We can easily identify a set of potential problems here. On what grounds are the experts entitled "experts"? Are there controversies involved in their *normative status*? If their skills are formalized and made explicit for the first time, can this *new terminology be used* in *practical situations* by others? Are there social problems related to *the creation of this terminology*? If these experts are replaced by a system, how do the *responsibility, work roles,* and *skill requirements* change for the persons depending on their expertise?

9.1 The Individualistic Fallacy within Computer Science

The inability to deal with such questions could be labeled *the individualistic fallacy*. It is caused by an attempt to use individualistic theories where social aspects are crucial to consider. In general, we believe that cognitive theories have played the general role, within computer science, to bring the human user into focus. However, as such they have had a too heavy task to fulfil. If cognitive science is used as *the* general theory about how technology is used in human

enterprises it will inevitably fail, even though it can make good contribution to that field.

In this paper, we have discussed how social issues are important to bring into focus in various ways, as a complement to the traditional approach in cognitive science. It is questionable, though, whether this is reasonable to do *within* cognitive science. How far should we extend that field? Cognitive science need not cover every aspect about human activities. Why should this discipline contain everything? We believe that current attempts to include "everything possible" in cognitive science, is related to the philosophical ambitions that has burdened this field from the start. As the ultimate philosophical theory of the mind, it has to include everything that is relevant to consider from some perspective or other.

A problem here, is that a social and an individualistic focus are not always compatible. Arguably you can not always extend individualistic theories to cover all relevant social aspects. There may be paradigmatic differences between the two concerns. Sometimes a focus on the group leads to an ignorance of the individual, and vice versa. If the social aspects, as listed above, are really crucial for the problem at hand, is it not more reasonable to simply look for theories within social sciences?

10 Summary

In this paper we have examined a set of criticisms of "the representation hypothesis" in traditional cognitive science. We have suggested a framework where we classify the criticism in three groups. We have also discussed some philosophical problems with the mental. We have argued that these problems can not be solved by simply extending the traditional approach to cover more phenomena, such as social, emotional, bodily, and situational aspects. Rather, these problems occur because we have too high expectations on what we can do with science in this context. Once we have realized this, we can give up our attempts to solve these questions within cognitive science, and we can start to appreciate the existing theories on a more practical basis. This, we have argued, must also include the traditional model of the mind as a manipulation of symbolic representations. Finally, we have discussed the "individualistic fallacy" within computer science, i.e., to use individualistic theories to address social problems.

11 References

Anderson J R (1980) *Cognitive Psychology and Its Implications*, New York, W. H. Freeman and Company

Clancey W (1991) Review of Rosenfield's 'The invention of Memory', *Artificial Intelligence*, 50

Collins H M (1990) *Artificial Experts*, Cambridge, Massachussets, the MIT Press

Dahlbom B and Janlert L-E (1988) *An artificiell värld* MDA-rapport 1988:1, Arbetsmiljöfonden

Dahlbom B (1993) Mind is Artificial, In: B Dahlbom (Ed.), *Dennett and His Criticis* Cambridge, USA, Blackwell

Dreyfus H L (1972) *What Computers Can't Do*, New York, Harper & Row Publisher

Dreyfus H L and Dreyfus S E (1986) *Mind over Machine*, New York, The Free Press

Malcolm N (1973) Thoughtless Brutes, *The Nature of Mind*, Rosentahl D M (Ed.), 1991 (pp. 454-461), Oxford University Press

Malcolm N (1982) The Relation of Language to Instinctive Behaviour, *Philosophical Investigations 5*, (pp. 3-22), Basic Blackwell Publisher Ltd

Malcolm N (1986) *Nothing is Hidden*, Blackwell

Suchman L (1987) *Plans and Situated Actions*, Cambridge, Cambridge University Press

Suchman L (1993) Reponse to Vera and Simon's Situated Action: A Symbolic Interpretation, *Cognitive Science*, 17 (pp. 71-76)

Svensson G (1991) *Wittgenstein - Om kunskap och visshet*, Stockholm, Thales

Winograd T and Flores F (1986) *Understanding Computers and Cognition*, Massachusetts, Addison-Wesley Publishing Company Inc.

Wittgenstein L (1953) *Filosofiska undersökningar*, Stockholm, Bokförlaget Thales, 1992

Wittgenstein L (1969) *Om visshet*, Stockholm, Lagerblads tryckeri, 1992