http://www.diva-portal.org

Postprint

This is the accepted version of a paper presented at *2nd International Conference on Agents and Artificial Intelligence (ICAART 2010), Valencia, Spain, January 22-24, 2010*.

Citation for the original published paper:

Billing, E. (2010)
Cognitive Perspectives on Robot Behavior.
In: Joaquim Filipe, Ana Fred and Bernadette Sharp (ed.), *Proceedings of the 2nd International Conference on Agents and Artificial Intelligence: Volume 2* (pp. 373-382). SciTePress
https://doi.org/10.5220/0002782103730382

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-12141

COGNITIVE PERSPECTIVES ON ROBOT BEHAVIOR*

Erik A. Billing
Department of Computing Science, Umeå University, Umeå, Sweden
billing@cs.umu.se

Keywords: Behavior based control, Cognitive artificial intelligence, Distributed cognition, Ontology, Reactive robotics, Sensory-motor coordination, Situated action.

Abstract: A growing body of research within the field of intelligent robotics argues for a view of intelligence drastically different from classical artificial intelligence and cognitive science. The holistic and embodied ideas expressed by this research promote the view that intelligence is an emergent phenomenon. Similar perspectives, where numerous interactions within the system lead to emergent properties and cognitive abilities beyond that of the individual parts, can be found within many scientific fields. With the goal of understanding how behavior may be represented in robots, the present review tries to grasp what this notion of emergence really means and compare it with a selection of theories developed for analysis of human cognition, including the extended mind, distributed cognition and situated action. These theories reveal a view of intelligence where common notions of objects, goals, language and reasoning have to be rethought. A view where behavior, as well as the agent as such, is defined by the observer rather than given by their nature. Structures in the environment emerge by interaction rather than recognized. In such a view, the fundamental question is how emergent systems appear and develop, and how they may be controlled.

1 INTRODUCTION

During the last decades, intelligent robotics has drawn towards a pragmatic view where no single design philosophy is clearly dominating. On the one hand, low level interaction with the world is often implemented with a reactive design philosophy inspired by Rodney Brooks' work, (Brooks, 1986; Brooks, 1990; Brooks, 1991a; Brooks, 1991b). On the other hand, classical AI-elements such as cartographers and planners are common modules for the high level control. Simultaneously, increasing system size and complexity raises requirements on well structured and modular system designs. Colored by an object-oriented programming approach, the system behavior is implemented through composition of modules. This kind of systems is commonly referred to as hybrid architectures. (Gowdy, 2000; Murphy, 2000; Doherty et al., 2004)

In a wider perspective hybrid systems propose a view of intelligence where simple behavior, like walking or grasping objects are typically reactive, while more complex tasks, like choosing a path or selecting objects are products of reasoning upon internal representation. "The robot can think in terms of a closed world, while it acts in an open world", (Murphy, 2000).

This view is not totally distant from the one proposed by modern cognitive science. The information processing model is still dominant for describing high level cognition (Stillings et al., 1995), while more reactive models have become popular for describing lower levels of control, especially within cognitive neuroscience (Shea and Wulf, 1995; Kaiser and Dillmann, 1996).

Even though hybrid architectures are today clearly dominating the field of intelligent robotics, there are several alternatives. A fundamentally different standpoint is taken by researchers proposing an embodied and holistic approach, (Mataric̀, 1997; Pfeifer and Scheier, 1997; Pfeifer and Scheier, 2001; Nicolescu, 2003). As these theories enforce concepts of distributed and emergent behavior, the present work is an attempt to analyze these notions of emergence actually mean. Similar ideas can be found within a vari-

*Parts of this text also appear as a technical report: Billing, E. A. (2007). Representing Behavior - Distributed theories in a context of robotics, *UMINF 07.25*, Department of Computing Science, Umeå University, Sweden.

ety of fields, including the *extended mind* (Clark and Chalmers, 1998), *distributed cognition* (Hutchins, 1995) and *situated action* (Suchman, 1987). All these theories are, in a general sense, studies of behavior beyond that of the individual, in groups or in interaction with artifacts. As such, these theories provide new perspectives on what it is we are in fact trying to achieve when building intelligent robots, and how we should get there.

1.1 Differences Between the Reactive and Deliberative Views

The reactive view grew during the 1980 as a reaction against classical artificial intelligence and cognitive science, and was in many ways a step back towards behavioristic ideas, (Braitenberg, 1986; Brooks, 1986; Georgeff and Lansky, 1987; Maes, 1991). The early reactive trend argued strongly against representations, but due to the obvious limitations of such an attitude, later work within the reactive field incorporate representations, but of a different type than the ones typically found within deliberating systems.

Deliberative architectures implement a *domain ontology*, that is, a definition of what things that exist in the world, but without a precise description of their properties and interrelations, (Russell and Norvig, 1995). This corresponds to a reductionist perspective also found within cognitive science and classical artificial intelligence.

Reactive systems are instead defined by a *low-level specification* that corresponds to the inputs and outputs of the system, referred to as the *sensory-motor space* (Pfeifer and Scheier, 1997). I here refer to a quite large variety of approaches, including the *subsumption architecture* (Brooks, 1986; Brooks, 1991b), *behavior based systems* (Mataric̀, 1997; Arkin, 1998; Nicolescu, 2003) and *sensory-motor coordination* (Pfeifer and Scheier, 2001; II and Campbell, 2003; Bovet and Pfeifer, 2005). Without claiming that all these approaches are one and the same, I use the term *reactive* as a common notion for these approaches proposing an embodied and holistic view.

The low-level specification defines the sensory-motor space as an entity which is related to the external world via a physical sensor or actuator on the robot. For a simple robot with eight proximity sensors and two independently controlled wheels, the sensory-motor space is ten-dimensional where each dimension corresponds to one sensor or motor. Many sensors provide multi-dimensional data. For example, a camera with a resolution of 100x100 pixels would increase the sensory-motor space with 10 000 new dimensions, where each pixel in the camera image corresponds to one dimension in the sensory-motor space. Cameras and other complex sensors could also be viewed as providing a single, complex, dimension in the sensory-motor space, but the amount of pre-processing or interpretation of data is always very limited in a low-level specification implement.

Pfeifer and Scheier (Pfeifer and Scheier, 1997) point out that a system using a low-level specification has a much larger input space than deliberative systems specified by a domain ontology, which also allows much greater complexity. Another interesting difference lies in information content. On the one hand, each dimension in the input space of a deliberative system is fairly informative. It could be the horizontal position of the robot on a map or the height of an object in front of the robot. On the other hand, most dimensions in the sensory-motor space are essentially meaningless if not viewed in the context of other dimensions. A single pixel in an image says very little about the content of the scene when that pixel is viewed alone, but in the context of the other pixels, it may be very informative. One could easily argue that such a large and complex sensory-motor space is the result of an ill chosen representation. With no doubt it is much easier to create readable representations using a deliberative approach, where sensor data have been processed so that it much better reflects *our own* understanding of what is going on.

Even though this criticism is correct and important, the sensory-motor space should not be understood as an unprocessed version of objects and other aspects in the world, but the representation of something else. The original argument against representations found in early reactive research has the last decade been replaced with a more accepting attitude towards representations, but representations of behavior rather than representations of the world. According to Pfeifer and Scheier, many things can be solved in a much simpler and more robust way without the use of high-level percepts. In general, the sensory-motor space appears to be a more suitable frame for representations than the kind of world models found in classical AI. (Pfeifer and Scheier, 1997; Pfeifer and Scheier, 2001; Dawson, 2002)

Low-level specifications have the great advantage that each dimension in the sensory-motor space is directly mapped to the corresponding sensors or motors, while the inputs to a deliberative system, such as position and size of objects, are often very hard to acquire. By assuming the necessity of high-level percepts we impose our own *frame of reference* upon the agent. Our notions of objects and states in the world are for

sure handy when reflecting upon an agent's behavior, but may not be necessary, or even desirable, when performing the same acts.

The principle of the frame of reference may be illustrated through the parable with the ant, presented by Herbert A. Simon, (Simon, 1969). Imagine an ant making its way over the beach, and that the way it chose was traced. When observing all the twists and turns the ant made, one may be tempted to infer a fairly complex internal navigation process. However, the complexity of the path may not be the result of the complexity of the ant, but the result of interaction between a relatively simple control system, and a complex environment.

Long before Brooks presented his ideas on reactive robotics (Brooks, 1986; Brooks, 1990; Brooks, 1991b), it was shown that complex behavior could emerge from simple systems, for example through the Homeostat (Ashby, 1960) and Machina speculatrix (Walter, 1963). Furthermore, Braitenberg's Vehicles (Braitenberg, 1986) was one of the most important inspiration sources for Brooks's work.

This discussion constitutes a central part of the criticism against deliberative systems and the motivation for a reactive approach. However, since reactive systems do not define any ontology with meaningful inputs, many types of, typically sequential tasks, are very hard to represent in this manner. Even though several examples of reactive systems showing deliberative-like behaviors exist, for example *Toto* (Matarič, 1992) and the reactive categorization robot by Pfeifer and Scheier (Pfeifer and Scheier, 1997), both the systems and the task they solve are typically handcrafted, making them appear more as cute examples of clever design than solutions to a real problem.

The difference between reactive and deliberative systems has been described as the amount of computation performed at run-time, (Matarič, 1997). A reactive control system can be derived from a planner, by computing all possible plans off-line beforehand, and in this way create a universal plan (Schoppers, 1987).

This argument about on-line computation beautifully points out how similar the two approaches of reactive and deliberative control may be. Still, when proposing the reactive approach, Rodney Brooks pointed out a number of behavioral differences to classical deliberative systems: "robots should be simple in nature but flexible in behavior, capable of acting autonomously over long periods of time in uncertain, noisy, realistic, and changing worlds", (Brooks, 1986). So if a reactive controller is merely a precomputed plan, why these differences in behavior?

One critical issue is speed. Brooks often points out the importance of real-time response and that the cheap design of reactive systems allows much faster connections between sensors and actuators than the deliberative planners, (Brooks, 1990). Even though this was an important point in the early nineties, the last years' increase in computational power allows continuous re-planning within a reactive time frame, (Dawson, 2002).

Another reason may be that reactive controllers are typically not derived from planners. Rather, reactive controllers are handcrafted solutions specialized for a certain type of robot. Achieving a specific complex behavior in a reactive manner can be a challenge, which may be one important reason for the limited success of reactive systems in solving more complex, sequential tasks (Nicolescu, 2003). Taking Matarić's point about run-time computation into account, the reactive approach still does not propose a clear way to achieve a desired controller; it only shows that the deliberative part can be removed when intelligence has been compiled into reactive decision rules.

Hybrid systems do obstacle avoidance using reactive controllers not because re-planning is computationally heavy, but because re-planning is difficult to implement. Even though one could imagine a planner generating exactly the same behavior as one of Braitenberg's vehicles avoiding obstacles, the structure of such a planner would probably be much more complicated than the corresponding controller formulated in reactive terms. This may in fact, at least from an engineer's point of view, be the most suitable distinction between the reactive and deliberative perspectives. It appears that behaviors like obstacle avoidance and corridor following is easily formulated in reactive terms, while selecting a suitable path from a known map is better formulated using a planner. Other things, actually most things, are too hard to manually design using any of these two approaches.

1.2 Emergence of Behavior

As mentioned in the previous section, supporters of the reactive approach freely admit that the implementation of high-level deliberative-like skills in reactive systems is very difficult, (Pfeifer and Scheier, 2001; Nicolescu, 2003). The route to success is often said to be emergence, (Maes, 1990; Matarič, 1997; Pfeifer and Scheier, 1997). But what exactly does this mean?

The term *emergent* is commonly described as *something that is more than the sum of its parts*, but apart from that it is in fact hard to arrive at a definition suitable for all uses of the term, (Corning, 2002). Within the field of intelligent robotics, emergence is used to point out that a robot's behavior is not explicitly defined in the controller, but something that ap-

pears in the interaction between the robot and its environment. Pfeifer and Scheier (Pfeifer and Scheier, 2001) proposes a number of design principles for autonomous robots. The critical points are shortly summarized below.

1. Behavior should emerge out of a large number of parallel, loosely coupled processes.
2. Intelligence is to be conceived as sensory-motor coordination, i.e., the sensory-motor space serves as a structure for all representations, including the categorization and memory.
3. The system should employ a *cheap* design and exploit the physics of its ecological niche.
4. The system must be redundant.
5. The system should employ the principles of self-organization.

Not surprisingly, those principles are well aligned with those found in literature discussing emergence, (Corning, 2002; Flake, 1998). Consequently, modern reactive architectures should constitute a good approach for design of systems showing emergent properties, but this is yet far from a unified theory on which robotics architectures could be built, (Heylighen et al., 2004). Before a theory of emergent behavior could actually be used, much more has to be understood about the theoretical properties of emergence, but such an analysis is seldom found in robotic literature.

1.3 Criticism Against the Hybrid View

After looking a bit closer at the principles of the reactive and deliberative approaches, the philosophy behind hybrid approaches seems to be much closer to the latter. Hybrid systems clearly align to a reductionist view, enforcing the importance of system modularity and hierarchical structures. The promoters of hybrid systems motivate this design effort in completely different ways than the supporters of reactive systems argue for a holistic perspective. Obviously, from an engineering perspective, it is very important to be able to build larger systems in some kind of modules, so that each part can be tested and refined separately. While this strongly contradicts the holistic perspective, reactive supporters have no solution to, and are generally not interested in, these issues.

So what exactly are the problems with combining the reactive in the small, and deliberative in the large? Since the hybrid approach is so wide and generally friendly towards anything that works, it is hard to actually say something about these systems which truly applies to all of them. Still, some common criticism has been raised against the hierarchical approach, especially from the field of embodied cognitive science. The core issues are summarized bellow.

• Even though hybrid systems adopt an embodied view for interaction with the world, they still define a domain ontology and are consequently bound by the limitations of this approach. One critical point of the reactive approach is that no concepts or symbols should be pre-defined. This point is lost when hybrid systems use reactivity as an interface to the world rather than the source of intelligence. (Pfeifer and Scheier, 2001)

• The kind of information produced by the reactive layer in a hierarchical system is often fundamentally different from that required by the deliberative subsystems, making it hard to design suitable interaction between the two layers. For this reason, the sensing part in the deliberative layer is often designed in a non-reactive way, reintroducing the problem of how objects, and concepts in general, should be recognized in complex and noisy data. (Pfeifer and Scheier, 2001; Dawson, 2002)

• While a critical aspect of modularity is to be able to test and control the function of each module before inserting it in the complete system, one important goal of reactive approaches is to achieve emergent properties which by definition do not appear in the modules alone. Even though hybrid systems successfully employ simpler reactive behavior, they do not leave room for emergent properties. (Pfeifer and Scheier, 2001; Brugali and Salvaneschi, 2006)

2 AN EXTENDED PERSPECTIVE

The fundamental differences between modern approaches within intelligent robotics have now been outlined. The rest of this paper present a number of different views on cognition, and apply them in a context of intelligent robotics. These views will make deliberative and reactive perspectives appear less like the two extremes, and more like one dimension within the multi-dimensional study of cognition and behavior.

Cognitive science has received significant amounts of criticism for its undivided focus on the individual, where a proper analysis of the social aspects of interaction is missing, (Greeno, 1993; Heylighen et al., 2004; Hutchins, 1995; Suchman, 1987). Similar criticism has been raised towards classical AI and deliberative robotics, but to some extent it also applies to reactive systems. Both deliberative and reactive approaches share a view of the single agent as one conceived unit, interacting

with the outside world through input and output interfaces. Even though this might seam like a safe assumption for many systems like humans, animals, computers and robots, I will in this chapter present a couple of theories where the object of analysis is changed to incorporate fundamentally different types of *cognitive systems*. As will be illustrated, many aspects of these systems are strikingly similar to the architectures proposed within robotics.

2.1 The Extended Mind

Clark and Chalmers (Clark and Chalmers, 1998) describe two people, Inga and Otto, who are both going to visit the Museum of Modern Art which lies in the 53rd street. Inga heard from a friend that there is an exhibition at the Museum of Modern Art and she decides to go there. She looks up the address and remembers it. She forms the belief that the Museum of Modern Art is on 53rd street. But now consider Otto who has Alzheimer's disease and instead of remembering the address writes it down in his notebook. Clark and Chalmers argue that Inga and Otto in principle have the same belief, even though parts of Otto's belief in a very strong sense are outside his body. This is an example of *the extended mind*.

Interestingly, Otto's behavior may easily be described in reactive terms. Otto changes the environment, his notebook, in a manner which later will lead him to the correct address. In contrast, Inga's behavior is a classical example which can't be described within a purely reactive architecture. Still, Clark and Chalmers point out that Otto's and Inga's cognitive processes are essentially the same. I would argue that the reason why we usually view Inga's and Otto's cognitive processes as quite different is our usual concept of an individual. If we chose to view the person as an entity enclosed by the skin, the use of a notebook as memory is very different from the use of nerve structures for the same purpose. But, if we instead follow Clark and Chalmers' argument and widen our notion of a person to include the notebook, the two types of memory appear very similar.

This point could also be illustrated by a computer. What exactly is a computer? Most people would probably say that it's the screen, the key-board and mouse, and of course the box on the floor which you connect all the cables to. If one uses a Memory Stick to store things on, that is not a part of the computer, but a different object. Still, technically speaking, the Memory Stick, when connected, is very similar to the hard drive within the computer. The Memory Stick might, just as Otto's notebook, have lower storage capacity, a bit slower access speed, and not always be available. Still, it fills the same function as the internal memory. Admittedly, our common notion of a person, where the notebook is not included, is very convenient, but we should be aware that it is merely a convention.

This discussion opens up the notion of an agent. We choose to see one agent as separated from its environment not because it *is* different from the environment, but because it, from our perspective, is convenient to view it like that. This does not imply that the notion of agents and objects is totally arbitrary, but neither is it totally predefined.

I mentioned earlier Pfeifer and Scheier's notion of *frame of reference*, pointing out that an agent's behavior is always seen from an observer's perspective. The view presented here takes one step further by saying that even the notion of agent is dependent of the observer. This distinction is important since Pfeifer and Scheier strongly argue towards representations within a sensory-motor space. If the agent is an entity created by the observer, so are sensors and motors, and with these, the sensory-motor space.

Following this discussion, the notion of an agent should be able to divide into smaller, sub-agents, with different sensors and motors. The functions of these sub-agents may differ drastically from the function of the combined agent, just like Otto and his notebook can do more things together, than neither of them can do separately. Consequently, the question of how behavior is represented is transformed into how Otto figures out that he should use a notebook? Or more generally: How does *purposeful* emergent behavior among agents appear?

2.2 A Universe of Possibilities

One of the inspiration sources to Clark and Chalmers' work came from *distributed cognition*, (Hutchins, 1995). Hutchins points out the importance of viewing cognitive processes as something that goes on both in the environment and within the individual, but in contrast to Clark and Chalmers, Hutchins' focus lies on group level dynamics. In this context, the agent, or the *cognitive system*, is expanded not only to incorporate one person and his tools, but many people and artifacts in cooperation.

Hutchins takes a few steps further than Clark and Chalmers by not only proposing an extended view, but also using it to analyze systems. Distributed cognition has been applied to many systems, including ship navigation (Hutchins, 1995), human-computer interaction (Hollan et al., 2000), various aspects of airplane control (Hutchins and Klausen, 1996; Hutchins and Holder, 2000; Hutchins et al., 2002) and more re-

cently clinical systems (Galliers et al., 2006). While distributed cognition as used in these examples have no apparent application to robotics, the result of this research might still shed some light on what we want to achieve when designing for purposeful emergent behavior.

From a deliberative perspective, a large and difficult problem is to recognize objects and their properties in complex and noisy data. From a reactive view, the same problem is instead described as how to arrive with suitable emergent properties. And finally, form the perspective of distributed cognition, this problem is strongly related to the formation of interpretations within a group. Hutchins investigates the properties of interpretation formation on group level using constraint satisfaction networks, (Hutchins, 1995). The weights of the networks were arranged so that each network could arrive at only two stable states, or interpretations. One interpretation corresponds to the activation pattern 111000 while the other interpretation corresponds to 000111, see Figure 1.

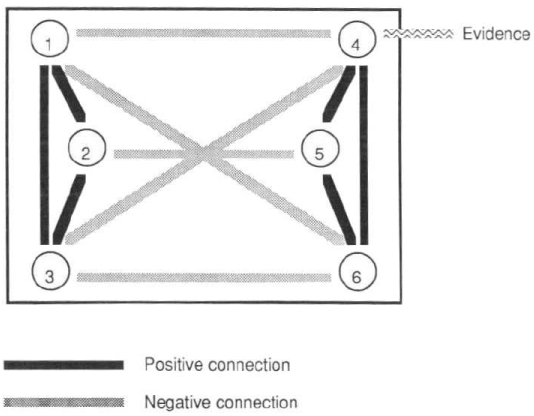

Figure 1: Constraint satisfaction network, (Hutchins, 1995, p. 244). Black and gray lines represent positive and negative connections, respectively. Strong activation in left side nodes will consequently inhibit activation in nodes to the right, and the other way, driving the network towards one of two stable states. Republished with permission.

The initial activation of the nodes is here viewed as *confirmation bias*. When the network is executed alone, it will always arrive with the interpretation closest to its initial state. However, when the networks are connected so that the activity of some nodes propagates to other nodes of a different network, their decision properties change due to interaction between the networks. As Hutchins puts it, this illustrates how interpretation of information changes due to the organization of the group. In a robotics context, this might be applied as having several reactive controllers creating a virtual "environment" for each other. More specifically, controllers do not only take input from sensors and send commands to actuators, but might also sense and act upon other controllers, drastically increasing the complexity of the system as a whole. The organization of such a system should be similar to the organizational properties of a group.

Hutchins shows that having this kind of organization, where multiple connected agents try to make their individual interpretations, produces a system that efficiently explores interpretation space. Furthermore, even after reaching a common interpretation, such a system is much more likely to re-evaluate the interpretation in case of new evidence. However, in case of too much interaction within the group, interpretation space is not explored properly. In contrast, too little interaction will result in that no single interpretation is achieved. Hutchins calls this *the fundamental tradeoff in cognitive ecology*.

This discussion is not only interesting as analysis of group behavior, but also as a way to understand interpretations within the individual. Here, the process of transforming sensor data into symbols with meaning is replaced with a continuous constraint satisfaction between sensors, actuators and internal states. In this view, we do not perceive what is in the environment; instead we are striving towards the closest stable state, pushed in one direction or another by changes in the environment.

There are mainly two advantages with this model of cognition, compared to the classical view of information processing. First, the interpretation always arrives from the current state of the system. Consequently, each interpretation is based on much more information than when sensor data is seen as a series of discrete readings which should be understood more or less separately. Secondly, when the notion of symbols is replaced with that of attractors, the number and meaning of these entities can be changed dynamically. This opens up the possibility for a solution to one key problem of deliberative processing; the creation of new symbols.

This view of cognition as the propagation of representational states across representational media might provide a new and powerful tool for understanding interpretation and decision making also in robots. However, even though Hutchins provides an extensive analysis of several existing distributed systems, a general understanding of how such a system appears and develops is still missing, (Heylighen et al., 2004).

2.3 Situated Action

When studying interactions between people, the analysis of language becomes one critical sub field. Clas-

sical cognitive science literature describes language as "a system that uses some physical signal (a sound, a gesture, a mark on paper) to express meaning", (Stillings et al., 1995). In other words, language is viewed as a communication channel where some meaning, i.e. an internal representation, is encoded into a physical signal using some grammar and then decoded by the listener into a similar internal representation. However intuitive this view may appear, it is not the only one. A fundamentally different perspective was presented in late eighties by Lucy A. Suchman under the name *situated action*, (Suchman, 1987).

Suchman's claim is that the traditional view of language includes several fundamental problems. One of the most important issues is the discussion around *shared knowledge*. If the cognitive view of language is correct, a speaker must not only encode the representation into words, but also take into account what the listener already knows. As Suchman puts it, this is equivalent with having a body of shared knowledge that we assume all individuals within our society to have. When speaking, only the specifics of the internal representation are transformed into words, leaving out everything covered by the shared knowledge.

To exemplify the problem, Suchman uses the result from an exercise assigned by Harold Garfinkel, (Suchman, 1987). Students were asked to write down a description of a simple conversation. On the left hand side of the paper the students should write what was said, and on the right hand side what they understood from the conversation. While the first part of the assignment was of course easy, the second part seemed to grow without limit. Many students asked how much they were supposed to write, and when Garfinkel imposed accuracy, clarity and distinctness the students finally gave up with the complaint that the task was impossible. The point here is not that the body of shared knowledge is too large to write down on a paper, but that the task resulted in a continually growing horizon of understandings to be accounted for.

The assignment, it turned out, was not to describe some existing content, but to generate it. As such, it is an endless task. The students' failure suggests not that they gave up too soon, but that what they were assigned to do was not what the participants in the conversation themselves did in order to achieve shared understanding. (Suchman, 1987).

Even though there might be several other ways to explain the result from Garfinkel's assignment, Suchman's point is striking. If knowledge is not preexisting to language as much as it is generated by it, it puts our understanding of internal representations in a fundamentally different light. The meaning of a spoken phrase does not appear to exist in any stronger sense than obstacles exists for one of Braitenberg's vehicles.

Situated action is not at all limited to analysis of language. In fact, situated action tries to unify all kinds of behavior where language is seen as a very specialized sub field. In such a view, spoken words has the same relation to semantics as an agent's actions has to intentions. However, it should be remembered that Suchman's work is presented as a theory of human-computer interaction rather than a theory of behavior or intelligence.

If language and other complex behavior do not begin with an internal representation or intent, how is it produced? The view proposed by Suchman begins in the context of the agent: "every course of action depends in essential ways upon its material and social circumstances," (Suchman, 1987). The circumstances or *situation of actions* can, at least in a context of intelligibility, be defined as "the full range of resources that the actor has available to convey the significance of his or her own actions, and to interpret the actions of others," (Suchman, 1987). This could be interpreted as if the world is understood in terms of actions. The fact that we know how to walk makes us really good at recognizing such behavior. In the domain of human-computer interaction, the same argument implies that our understanding of a computer is represented in terms of what we can do with it, not as a structural model of the computer as such. As a consequence, a selection of the best path towards a desired goal is not dependent on a representation of roads, but on the availability of the actions for turning left or right.

Furthermore, the *goal* of situated action is not represented in any other way than as preferences to some actions given a specific situation. Our common understanding of plans and goals is in this context nothing but a way to reflect on past events. As Suchman points out, a declaration of intent generally says very little about the precise actions to follow, it is the obscurity of plans that makes them so useful for everyday communication (Suchman, 1987).

This discussion puts not only notions of intentions and plans in a secondary position, but conscious thoughts in general appear to be less the driving force behind action than an artifact of our reasoning about action. Seen in the robotics context, deliberative processes should in a very strict sense be emerging from lower levels of interaction, not something predefined that supervises the lower levels.

One interesting implication of these theories is that observed sensor data bears a very loose connection to its semantic content. The interpretation is *cre-*

ated by the observer in interaction with the data rather than *extracted* from the observed data. The creation of an interpretation is in this view more about generating information, than processing it.

3 DEVELOPMENT AND DESIGN

The theories presented above depict a perspective of intelligence where cognitive ability emerges out of interactions between multiple parts of an agent. The agent is very loosely defined as a cognitive system, i.e., a large number of physically and/or socially distributed entities which interact and in this way achieve something more than any of them could do alone. More explicitly, intelligent behavior of a cognitive system is produced from entities which are totally unaware of the dynamics of that system as a whole.

In such a view, elements in the world are never explicitly represented, but appear in terms of possibilities for situated actions. Information does not flow from inputs to outputs, but back and forth through numerous representational states, coordinating sensors and actuators rather than controlling them. The meaning of actions, symbols or data in general is achieved through interaction among elements, not given by a grammar. Conscious processes are not the fundamental foundation for intelligent behavior, but its fundamental phenomenon. The question still remaining is: *How does one create a system based on the principles presented above?*

3.1 Evolution of Self-organization

Emergent properties which in the previous discussion so gracefully are said to explain intelligence do, from an engineering perspective, often cause more problems than they solve. What is seldom mentioned is that guidelines like those presented by Pfeifer and Scheier, (Pfeifer and Scheier, 2001) only address half the question of emergent behavior. We are normally not interested in just any emergent behavior, but specifically in those emergent effects which fulfill the task for which the robot is designed. This may prove to be much more difficult to achieve than just emergence in general.

There are a number of theories approaching this problem. The most frequently mentioned within robotics is self organization. The principle of self organization means that the system spontaneously develops functional structure through numerous interactions between its parts. The basic mechanism behind this structure is mutual benefit, symbiosis. Parts in the system will continue to rearrange until both find a relative state which is satisfactory. A frequently used interaction pathway will grow stronger while rarely used pathways will weaken or disappear. The mechanisms controlling what is satisfactory will as a consequence have direct influence on the emergent properties of the system as a whole. (Heylighen and Gershenson, 2003)

While the principle of self-organization provides a clearer understanding of the mechanisms controlling emergence, it still does not explain how useful behavior emerges. The fact that favorable interactions are reinforced on the micro level will certainly not lead directly to favorable behavior on the macro level.

For natural systems the obvious answer is evolution. The fundamental mechanism of natural selection will, given enough time, lead to a solution. However, this gives us little hope for training robots. The problem space for a robot acting in the real world grows extremely fast. Allowing a robot to try out a population of randomly chosen behaviors will, even for very simple problems, most likely never lead to a solution. Consequently, the evolutionary process won't work since nothing can be reinforced.

Interestingly, a robot acting in the real world is, by definition, in the same situation as many biological systems, which obviously have evolved. The discussion above makes an evolutionary explanation to the problem of grabbing and moving objects appear highly unlikely. Even when including the billions of years preceding the human era, the chance of combining all the biological structures required for object manipulation to work appears very small. And yet evolution has given us a wonderful tool in the form of the hand, and the neural structures underlying its control.

The explanation for our highly flexible and dexterous ability to manipulate objects is of course that it did not evolve from nothing. As such, the human hand is not an optimal solution, nor is it anything close to optimal. Instead is it a result of what came before it. Small incremental changes of the mammal front legs, which at each stage were reinforced through natural selection, have eventually led to the human arm and hand. (Wolfram, 2002)

Why this divergent discussion about evolution? The manipulator of a robot has to be designed. We are simply interested in teaching the robot to use it, given a fairly short period of time. The point of this sidestep into evolution is that the physical shape of the human hand did not evolve alone, but together with the neural system controlling it. The human child is born with a large amount of basic reflexes, which are all fairly simple. In robotic terms, we would probably call them purely reactive behaviors. (Thelen and

Bates, 2003; Grupen, 2006)

Some of these innate behaviors are certainly critical for the survival of the infant, such as the grip reflex or the sucking reflex. But many other reflexes do not have an obvious purpose, such as the *asymmetric tonic neck reflex*, *landau reflex* or the *galant reflex*, (Grupen, 2006). Instead, reflexes like these seem to play a key role for learning. As mentioned above, the child is born with a set of reflexes. These basic reflexes are, during the first four years, gradually replaced by new, more complicated behaviors. The child seams to learn through an evolutionary process of behavioral development. New behaviors appear as a modification or combination of more basic behaviors, while other behaviors disappear. In theoretical terms, this incremental development allows the problem space to remain small even as the problems grow more complicated. The full space of possible solutions to the problem is never searched, but only the parts covered by previous knowledge. Sometimes, the small changes to the underlying control structure result in drastic changes in behavior, which we see in the child as the establishment of new behaviors. This kind of incremental development process should favor robust behavior rather than optimal. A behavior which succeeds under many circumstances simply has much greater chance of survival than a perfect solution only succeeding under very special circumstances.

4 CONCLUSIONS

Through out this review there has been a theme of emergent behavior. Notions of objects in the world, goals and concepts in general are said to emerge out of simpler parts. The literature reviewed here frequently points out several aspects critical for a system to show emergent properties. However, it has been much harder to find clear theories of how to control these emergent properties. In fact, one important property of emergence seems to be that it can not be controlled in a supervising way. Without a proper theory of how to arrive at useful emergent properties, the argument that behavior should emerge is very much like saying that we do not know. It is generally admitted that distributed and emergent control systems for robots are very hard both to obtain and control. As such, these approaches do not seem less problematic than their counterparts within classical AI.

Nevertheless, the concepts presented in this review open a new frame for representation of behavior. A troubling and yet thrilling aspect of these theories is that they span over an enormous theoretical horizon. Some fundamental problems are solved, many new are introduced, but just viewing the problem from a different perspective might get us closer to a general understanding: An understanding of what intelligence is and how it can be created.

ACKNOWLEDGEMENTS

I thank Lars-Erik Janlert and Thomas Hellström at the Department of Computing Science, Umeå University for valuable input to this work.

REFERENCES

Arkin, R. C. (1998). *Behaviour-Based Robotics*. MIT Press.

Ashby, R. (1960). *Design for a brain; the origin of adaptive behavior*. Wiley, New York.

Bovet, S. and Pfeifer, R. (2005). Emergence of coherent behaviors from homogenous sensorimotor coupling. In *12th International Conference on Advanced Robotics*, pages 324 – 330.

Braitenberg, V. (1986). *Vehicles - Experiments in Synthetic Psychology*. The MIT Press.

Brooks, R. A. (1986). A robust layered control system for a mobile robot. In *IEEE Journal of Robotics and Automation RA-2*, volume 1, pages 14–23.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, 6:3–15.

Brooks, R. A. (1991a). Intelligence without reason. *Proceedings, 1991 Int. Joint Conf. on Artificial Intelligence*, pages 569–595.

Brooks, R. A. (1991b). New approaches to robotics. *Science*, 253(13):1227–1232.

Brugali, D. and Salvaneschi, P. (2006). Stable aspects in robot software development. *International Journal of Advanced Robotic Systems*, 3(1).

Clark, A. and Chalmers, D. J. (1998). The extended mind. *Analysis*, 58:10–23.

Corning, P. A. (2002). A venerable concept in search of a theory. *Complexity*, 7(6):18–30.

Dawson, M. R. W. (2002). From embodied cognitive science to synthetic psychology. In *Proceedings of the First IEEE International Conference on Cognitive Informatics (ICCI'02)*.

Doherty, P., Haslum, P., Heintz, F., Merz, T., and Persson, T. (2004). A distributed architecture for autonomous unmanned aerial vehicle experimentation. In *Proceedings of the 7th International Symposium on Distributed Autonomous Systems*.

Flake, G. W. (1998). *The Computational Beauty of Nature*. MIT Press, Cambridge, Massachusetts.

Galliers, J., Wilson, S., and Fone, J. (2006). A method for determining information flow breakdown in clinical systems. *Special issue of the International Journal of Medical Informatics*.

Georgeff, M. P. and Lansky, A. L. (1987). Reactive reasoning and planning. In *AAAI*, pages 677–682.

Gowdy, J. (2000). *Emergent Architecture: A Case Study for Outdoor Mobile Robots*. PhD thesis, The Robotics Institute. Carnegie, Carnegie Mellon University.

Greeno, J. G. (1993). Special issue on situated action. In *Cognitive Science*, volume 17, pages 1–147. Ablex Publishing Corporation, Norwood, New Jersey.

Grupen, R. (2006). The developmental organization of robot behavior. Oral presentation during the 6th International UJI Robotics School.

Heylighen, F. and Gershenson, C. (2003). The meaning of self-organization in computing. In *IEEE Intelligent Systems*.

Heylighen, F., Heath, M., and Overwalle, F. V. (2004). The emergence of distributed cognition: a conceptual framework. In *Collective Intentionality IV*, Siena, Italy.

Hollan, J., Hutchins, E., and Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.*, 7(2):174–196.

Hutchins, E. (1995). *Cognition in the Wild*. MIT Press, Cambridge, Massachusetts.

Hutchins, E., E, B., Holder, R., and P, A. (2002). Culture and flight deck operations. Prepared for the Boeing Company.

Hutchins, E. and Holder, B. (2000). Conceptual models for understanding an encounter with a mountain wave. In *HCI-Aero 2000*, Toulouse, France.

Hutchins, E. and Klausen, T. (1996). Distributed cognition in an airline cockpit. *Cognition and communication at work. Y. E. a. D. Middleton. New York, Cambridge University Press*, pages 15–34.

II, R. A. P. and Campbell, C. L. (2003). Robonaut task learning through teleoperation. In *Proceedings of the 2003 IEEE, International Conference on Robotics and Automation*, pages 23–27, Taipei, Taiwan.

Kaiser, M. and Dillmann, R. (1996). Building elementary robot skills from human demonstration. *International Symposium on Intelligent Robotics Systems*, 3:2700–2705.

Maes, P. (1990). Situated agents can have goals. *Robotics and Autonomous Systems*, 6:49–70.

Maes, P., editor (1991). *Designing Autonomous Agents*. MIT Press, Elsevier.

Matarič, M. J. (1992). Integration of representation into Goal-Driven Behavior-Based robots. In *IEEE Transactions on Robotics and Automation*, volume 8, pages 304–312.

Matarič, M. J. (1997). Behavior-Based control: Examples from navigation, learning, and group behavior. *Journal of Experimental and Theoretical Artificial Intelligence*, 9(2–3):323–336.

Murphy, R. R. (2000). *Introduction to AI Robotics*. MIT Press, Cambridge, Massachusetts.

Nicolescu, M. (2003). *A Framework for Learning from Demonstration, Generalization and Practice in Human-Robot Domains*. PhD thesis, University of Southern California.

Pfeifer, R. and Scheier, C. (1997). Sensory-motor coordination: the metaphor and beyond. *Robotics and Autonomous Systems*, 20(2):157–178.

Pfeifer, R. and Scheier, C. (2001). *Understanding Intelligence*. MIT Press. Cambrage, Massachusetts.

Russell, S. and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall, NJ.

Schoppers, M. (1987). Universal plans for reactive robots in unpredictable demains. In *IJCAI-87:*, pages 1039–1046.

Shea, C. H. and Wulf, G. (1995). Schema theory - a critical appraisal and reevaluation. *Journal of Motor Behavior*.

Simon, H. A. (1969). *The Sciences of the Artificial*. MIT Press, Cambridge, Massachusetts.

Stillings, N. A., Weisler, S. E., Chase, C. H., Feinstein, M. H., Garfield, J. L., and Rissland, E. L. (1995). *Cognitive Science*. MIT Press, Cambridge, Massachusetts.

Suchman, L. A. (1987). *Plans and Situated Actions*. PhD thesis, Intelligent Systems Laboratory, Xerox Palo Alto Research Center, USA.

Thelen, E. and Bates, E. (2003). Connectionism and dynamic systems: Are they really different? *Developmental Science*, 6(4):378–391.

Walter, W. (1963). *The Living Brain*. Norton & Co., New York.

Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media, 1 edition.