ELSEVIER

Cognitive Systems Research 3 (2002) 339–348

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

When is a cognitive system embodied?

Action editor: Tom Ziemke

Alexander Riegler

CLEA, Vrije Universiteit Brussel, Krijgskundestraat 33, B-1160 Brussels, Belgium

Received 31 May 2001; accepted 30 October 2001

Abstract

For cognitive systems, embodiment appears to be of crucial importance. Unfortunately, nobody seems to be able to define embodiment in a way that would prevent it from also covering its trivial interpretations such as mere situatedness in complex environments. The paper focuses on the definition of embodiment, especially whether physical embodiment is necessary and/or sufficient for cognitive systems. Cognition is characterized as a continuous complex process rather than ahistorical logical capability. Furthermore, the paper investigates the relationship between cognitive embodiment and the issues of understanding, representation and task specification.

Keywords: Complexity; Constructivism; Design; Embeddedness; Representation; Teleonomy; Understanding

1. Introduction

We all are living beings. As such, if we look around we perceive a world full of different shapes, colors, sounds, smells, etc. All these modalities are not in merely chaotic disorder. By using our cognitive abilities, we find sense in the world. We can identify objects, and distinguish them from one another. We also can manipulate things, thus changing relationships between them. Making sense of the world helps us to survive. This all seems very self-evident and trivial to us. But is it? We have not had most of our capabilities at birth. For a newborn baby, almost nothing is obvious. There is an ongoing (and actually never-ending) process of cognitive development which makes us what we are.

But what about our pets, what about animals? Do they just experience a chaotic disorder of color and noise? Apparently they don't either. Observing animals, either as an amateur or an ethologist, reveals their remarkable capabilities. So is there a difference? Obviously, most animals do not talk to each other, and perhaps all have problems in solving mathematical equations. We conclude, there must be a gradual difference between humans and dogs, between cats and amoebas. One thing they all have in common: strategies to survive, or more precisely, to live. Strategies to be developed and executed need a cognitive apparatus of some sort. In other words, creatures survive in their environments by using their cognitive abilities which is in turned shaped by the interaction with the environment. As Rodney Brooks puts it, ''Intelligence is determined by the dynamics of interaction with the world'' (Brooks, 1991, p. 585). This is, in a nutshell, the motivation for developing the concept of embodiment, which entered cognitive science and related disciplines in the 1980s.

E-mail address: ariegler@vub.ac.be (A. Riegler).

1389-0417/02/$ – see front matter
PII: S1389-0417(02)00046-3

In this paper we will focus on the definition of embodiment, especially whether physical embodiment is a necessity for cognitive systems, and its relationship to the issues of understanding, representation and task specification. We will make use of the idea that cognition is a continuous complex process rather than an ahistorical logical capability. The goal is to come up with a clearer idea of the role of embodiment for cognitive systems.

2. Cognition

The complexity of cognition—defined in terms of behavioral repertoire that enables adequate compensation of perturbations from the environment—is different for different individuals: as we have seen in the Introduction, there is a large-scale between simplest forms of cognition in simple lifeforms and human-like cognition.[1] If we trust in the idea of evolution we have to ask for the mechanisms that evolved the wide range of different cognitive apparatus over time.

When we speak about evolution, we implicitly assume a gradual adaptation of species towards their environment. In 1941, Konrad Lorenz, for example, stated that the horse's hoof is a representation (*Abbild*) of the steppe, the body form of the dolphin is the incarnation of knowledge about laws of aerodynamics in water, etc. (Lorenz, 1982).[2] Likewise, cognitive capabilities seem to have emerged in resonance to the structure of the environment (resulting in 'anschauungsformen' in the terminology of Lorenz). We are intelligent because our environment has been challenging enough to select for the smart guys. The idea that the physical world influences the behavior of an agent (rather than being fed with instructions from the programmer) is commonly referred to as 'situatedness' (Pfeifer & Scheier, 1999). Embodied beings deal with the world, and their cognitive capabilities emerge out of this interaction.[3]

How can we conceive of the interaction between system and environment as the engine of cognitive complexification? Taking Lorenz's statement further, cognitive capabilities are also a reflection of the environment.[4] This idea leads directly to the proposal that Karl Popper labeled the 'bucket theory of mind' (1979). According to this view, ''there is nothing in our mind which has not entered through our senses''. Thus, animals and humans are considered cognitive buckets which get filled over time. While this concept has much common sense attractiveness, it quickly runs into problems such as the frame problem (Dennett, 1984). How do you feed your cognitive apparatus with the facts of the 'outside world'? How do you formally specify what changes in your environment and what remains constant? As each fact in the world is potentially connected with any other piece of fact we would need to update the content of our bucket each time there is a change in our environment. The wrong assumption here is that the world is a collection of facts that could be arbitrarily combined with each other. Even if we managed the combinatorial complexity, a question would remain: what is a *fact*? Entities in our perception don't come labeled (Franklin, 1995). If we look at a tree, we know that it is a tree—but how has this meaning emerged (cf. the symbol grounding problem, Harnad, 1990)?

[1] Several authors expressed the tight relationship between life and cognition. ''Living systems are cognitive systems, and living as a process is a process of cognition'' (Maturana & Varela, 1980, p.13). Adolf Heschl (1990) writes that both terms ''... are revealed as truly synonymous notions'' (p. 18). However, the title of his and John Stewart's (1996) papers, Life = Cognition, bears the problem that any cognitive system must be unreflectedly considered alive as well. My position is more of the order relation ''Life ⊂ Cognition'', i.e. artifacts can be cognitive without being a living system.

[2] 'Our categories and forms of perception, fixed prior to individual experience, are adapted to the external world for exactly the same reasons as the hoof of the horse is already adapted to the ground of the steppe before the horse is born and the fin of the fish is adapted to the water before the fish hatche' (Lorenz, 1982, pp. 124–125). In Lorenz (1973, p. 39) he states that ''*im* lebenden System eine *Abbildung* der realen Außenwelt entsteht [...] ein *Negativ* der Wirklichkeit [an *image* of the material world is built up *within* the organism ... a photographic *negative* of reality]''.

[3] Space does not allow for discussing *social* embodiment—being situated within a society and growing up there (cf. Dautenhahn, 1997). Since other individuals are part of one's world the social dimension is a sophistication rather than a contradiction of what is being said in the present paper.

[4] Cf. also the analogy of Herbert Simon (1969) referring to an ant walking along the beach. Simon argued that the complexity of the behavior is a reflection of the complexity in the environment.

Scientists aware of these tricky and annoying problems, which made artificial intelligence an unreachable goal in the past, proposed the concept of embodiment as a solution. It refers to the idea that ''intelligence cannot merely exist in the form of an abstract algorithm but requires a physical instantiation, a body'' (Pfeifer & Scheier, 1999). Through embodiment, symbols get physically grounded, and such meaning is defined through interaction with the world. However, the importance of physical embodiment has been questioned. Oren Etzioni (1993) criticizes that building robots which interact with the complexity of the 'real world' are not the exclusive road to embodied cognition. Operating systems such as UNIX too provide a sufficiently complex and hence challenging environment for cognitive agents. Quick, Dautenhahn,niv,and Roberts (1999) propose (non-optimal) information retrieval agents on the internet to be candidates of embodiment. Their 'phenomorph' program is based on the idea that its relationship to the internet is analogous to the one of the bacterium *Escherichia coli* to its environment. Stan Franklin (1997) discusses embodiment for a whole class of agents irrespective of the realm they inhabit. According to Franklin, intelligent systems ''must be embodied in the situated sense of being autonomous agents *structurally coupled* with their environment'' (Franklin, 1997, p. 500, my emphasis). These and other authors share the common view that embodiment does not necessarily mean physically embodied.

3. Structural coupling

In order to judge the importance of *physical* embodiment, let us turn to a key notion of Franklin's quote. What does it mean for a system to be *structurally coupled*? The answer requires to clarify what a system consists of. We may distinguish between the typical characteristics of a system defined through the relationships among its components, and the actual components that are involved in establishing these relationships. While it is clear that there is an arbitrary large number of possibilities how components can be arranged in order to fulfil certain relational criteria, this set of criteria itself must be invariable. Humberto Maturana and Fransisco Varela (Maturana & Varela, 1980) proposed to label the set of criteria which uniquely define a class of systems its *organization*. The actual components that fulfil the criteria are subsumed as the *structure* of the system in question. Clearly, there is a homomorphic relationship between organization and structure, as many different structures can establish a particular organization. In other words, the structure can vary without losing its constitutive character for the organization. These variations can be caused by perturbations to the system. Synonymously we can speak of structural deformations which occur when perturbations have an effect on the system in question. Let's think of a watch. We can exchange parts of the watch (modify its structure) without changing its functioning as a watch (its organization). Actually, this is what watchmakers are doing when they seek to repair a watch, namely exerting 'perturbations' to it.

Quick et al. (1999) provide us with a 'minimal' definition of embodiment which is based on Maturana and Varela's notions of organization, structure, and perturbation: ''A system X is embodied in an environment E if perturbatory channels exist between the two. That is, X is embodied in E if for every time t at which both X and E exist, some subset of E's possible states have the capacity to perturb X's state, and some subset of X's possible states have the capacity to perturb E's state''. This definition leads directly to the concept of structural coupling. If perturbations appear mutually between a system and its environment, the system is said to be structurally coupled. In the perspective of Quick et al., a system is embodied if it is structurally coupled with its environment. While this attempt to clarify the notion of embodiment is an important first step, it is at the same time an insufficient characterization. Of course, every system is in one sense or another structurally coupled with its environment. This applies to living creatures as well as to computer programs, since both are exposed to perturbations. (In the case of a computer program think of keystrokes as perturbations to a text-processing system.)

Clearly, structural coupling is a matter of mutual interactivity. ''A fly seen walking on a painting by Rembrandt does not interact with the painting by Rembrandt. The painting by Rembrandt exists only in the cultural space of human aesthetics, and its properties, as they define this cultural space, cannot

interplay with the properties of the walking fly'' (Maturana, 1980, p. 51). In agreement with the definition of Quick et al., we can state that a system has to be sensitive to perturbations in order to get structurally coupled with the other system which emanates the perturbations. In the example neither the fly nor the picture are sensitive to perturbations of the other. Of course, we *could* argue that there is some interaction going on at the molecular level, even in the case of a light-weighted fly that walks over a painting. However, these effects are dampened out.[5] It is important to note that perturbations may also arise in other domains than physical space. As the quote suggests, the effect a painting has on a human spectator lies in an intellectual–cultural rather than physical space.

Evidently, the fly did not 'understand' (or 'appreciate') the Rembrandt painting. Given the complexity of the 'real world' (within which fly and painting are embedded) we are suspicious that the ignorance of the fly has something to do with the way it handles 'overload' with potentially infinite information. As we will see, the richness of natural environments cannot be compared with those in artificial environments.

4. Embodied understanding and representation

Quantum physics aside for a moment, let us consider the world being a gigantic system with an (literally) astronomical number of states which interact deterministically. What science does is to build models that relate to the world system in a homomorphic way. This means that we design the model with appropriate combinations of states of the world system such that the state transitions in the model remain deterministic. Such is our 'view' of the world. After all, our cognitive capabilities are limited. Thus, we have to come up with models of limited complexity in order to manage them. In other words, models are our way to *understand* the world (Riegler, 1998). As engineers we implement these intellectual vehicles *back* into the world, for example as robots. Of course, what was an appropriate combination of states for our understanding need not be a proper combination for robots.

There is a parallel in artificial intelligence. Block worlds in the 1970s, such as Winograd's Shrdlu (1972), were a straightforward attempt to implement our simplified model-based understanding of the world. The number of entities was small, and so was the number of possible combinations of how those entities could be related to each other. The motivation for this approach is clear. In logical calculus, semantics defines meaning and truth in terms of an underlying model, ontology, or logical interpretation. Thus, if you consider causal action–effect relationships as rules of inference, objects as axioms, and facts as starting axioms, it becomes possible to treat the 'real world' in exactly the same way as a mathematical formalism. There is no problem to draw logical conclusions in such a system. Problem solving in the sense of Newell and Simon's means–ends view (e.g. Simon, 1969) is manageable since the number of intermediate steps is small.

However, if systems become bigger, the combinatorial explosion of possible links between entities grows faster than any computer or living cognitive system could ever compute. Feedback loops in the relational network create non-linear effects. Linearly decomposable blockworlds become non-linear *Dennett-situations*[6] in which logically operating robots are doomed to failure (Dennett, 1984). All computation is devoted to logical inferences so that there is no time for action in such scenarios.

These fundamental problems granted influence to critics of artificial intelligence. They questioned the representational paradigm according to which there is a more or less stable referential mapping between system and environment. Of course, how can logical symbols upon which the reasoning of such systems is based be related to anything else than other symbols? Critics doubted that AI systems were ever to understand the domain which they populate.

Noel Sharkey and Tom Ziemke (2000) mention the astonishing capabilities of the horse Clever Hans as an example of apparent understanding. The example also outlines the difference between self-con-

[5] Cf. also the idea of nearly-decomposable systems of Simon (1969) as referred to later in the paper.

[6] The dilemma from which a robot suffers which is given the task of retrieving a spare battery from a room with a bomb. Either it ignores the logical implications of its actions (thus risking to overlook the bomb) or it gets simply overwhelmed by the number of logical implications.

trolled systems and systems which are determined from the outside. Clever Hans could solve arithmetic tasks by tapping the solution with its hoof. As it turned out later, the animal solved problems by reacting to cues from the audience, which caused it to stop tapping at the right moment, rather than solving the problem by understanding. It is clear that for an animal human written numbers and arithmetical signs do not make sense. The horse is not embodied in the domain of arithmetic. Its behavior has meaning only for the spectators who are surprised to see a horse capable of doing mathematics.

Let's continue with this thought. It evidently leads us to the question: what is 'correct understanding'? A more human-centered example will shed some more light on this question and its possible answer. Richard Feynman (1985) introduced the notion of 'Cargo Cult Science'. Inhabitants of a fictive island in the South Sea had witnessed the support of goods by airplane during World War II. Of course they wanted this to happen again. So they started to create runways with fires along their sides; they set up a wooden hut for a man to sit in, with two wooden plates on his head as headphones, and bars of bamboo sticks looking like antennas. The form was perfect; everything looked the way it had been looking before. But, not surprising to us, it did not work; no plane ever landed. From the perspective of embodiment, the lack of understanding results from a lack of being embodied in the world of Western science and technology.

While the (intellectual) world of horses seems to be quite distant from that of humans, the story of Clever Hans and the analogy of Cargo Cult thinking are related to the same issue. The inability to understand results from the lack of synchronization (structural coupling) and, consequently, from the lack of embodiment. Again, understanding cannot arise from mechanically working off meaningless symbols which are designed as stable references to states in the world. Ernst von Glasersfeld (1983) provides us with an illustration which hints at a different explanation of 'understanding'. Suppose that, for the first time, you hear the word 'mermaid'. You are told that it is a hybrid creature between woman and fish. It is easy to construct a representation out of already known elements which are associated with 'woman' and 'fish', namely a composite which is a fish-tailed biped and, therefore, rather unlike the intended creature of the sea. You can go on reading stories about mermaids without getting into conflict with your deviant notion, unless you encounter a picture of a mermaid. Therefore one will modify the concept that is the subjective interpretation of the word only if some context forced him or her to do so.

Clearly, representation and understanding are system-centered rather than referential (Peschl & Riegler, 1999). Meaning arises as a result of relating a new piece of experience to the existing network of already made experiences rather than to entities in the world. The essential difference to Popper's bucket theory is that it is not a fact which is integrated in an existing universally valid semantic network of facts. What for one individual is a meaningful entity might go unnoticed for another. Also if two individuals process similar experiences it does not mean that these experiences are integrated in the same way. These issues have been extensively discussed in philosophy, especially within the epistemology of constructivism (e.g. Uexküll, 1934; Glasersfeld, 1995; Riegler, 2001a). However, we don't want to content ourselves with pointing at the constructive nature of knowledge and understanding. Rather we want to investigate possible accounts. As we'll see, embodiment plays a major role in them.

5. Experience on demand

Ulric Neisser's (1976) characterization of perception as a schemata-controlled 'information pickup' corresponds to the constructivist perspective. An organism's cognitive apparatus, i.e. schemata, determines the way it is looking at the environment. The schemata construct anticipations of what to expect and thus enable the organism to actually perceive the expected information. Without anticipation no 'information' (cf. also Riegler, 2001b). Neisser speaks of a 'perceptual cycle': 'the schema accepts information as it becomes available at sensory surfaces and is changed by that information; it directs movements and exploratory activities that make more information available, by which it is further modified' (p. 55). Therefore we have a mutual interplay between the cognitive apparatus and the information it retrieves. Note that 'information' only makes sense for the individual who integrates it

into the existing network of schemata. The network may undergo a modification due to the integration of the new experience. Obviously, this is reminiscent to Jean Piaget's (1954) notions of assimilation (pickup and integration of information) and accommodation (modification of the cognitive apparatus). We can even go one step further and point to the canalizing aspect of schemata-driven information pickup (Riegler, 2001b). This perspective of cognition reverses the information-processing paradigm. We can no longer speak of information input. No longer are organisms exposed to information overload as a result of *processing* the entirely available information which is filtered for relevant issues in order to control their behavior. Rather, we may conceive of 'perceptive interaction on demand' of the cognitive apparatus. I have labeled this Popperian 'searchlight view of mind' (1979) the 'constructivist–anticipatory principle' (e.g. Riegler, 1994).[7]

The global picture is that cognition acts independently of the environment. It merely requests confirmation for its ongoing dynamical functioning. This way, embodiment gets a whole new dimension. Neither is cognition cut off from its environment (like logical AI programs) nor is it at the mercy of its environment (like radical proponents of embodiment declare in statements like ''the world is its own best model'', Brooks, 1991). Instead, cognition works autonomously. It is *organizationally closed* (Maturana & Varela, 1980) in the sense that the nervous system is ''a closed network of interacting neurons such that any change in the state of relative activity of a collection of neurons leads to a change in the state of relative activity of other or the same collection of neurons'' (Winograd & Flores, 1986, p. 42). Structural coupling takes place at the demand of the cognitive apparatus.[8] In order to guarantee a proper functioning of cognition the process has to be 'synchronized' with the environment. It is practically impossible to explicitly design a cognitive artifact that fits perfectly in the appropriate niche.[9] 'Embedding' (situating) such an artifact in an environment is simply not enough. This situation compares to the attempt to exchange parts of a watch with arbitrary components at random, expecting that the parts would again contribute to the functioning of the watch. Rather, structures have to develop. In other words, the system has to be embodied in its environment. In this sense, the acting of a watchmaker can be seen as embodying new spare parts because the watchmaker knows about the organization of the watch. However, Giambattista Vico's 'Verum ipsum factum'[10] reminds us of the fact that human-made systems are far easier to understand than natural ones. So what is possible for a watchmaker (with regard to watches) is not by any means possible for AI scientists (with regard to cognitive artifacts). In the following we will more carefully develop an explanation why.

6. Embodiment of artifacts

In their attempt to marry Jakob von Uexküll's work with Maturana and Varela's (1980) concept of autopoiesis, Sharkey and Ziemke (2000) arrive at a conclusion that could be called ''the biochemical integrity of living systems''. This integrity distinguishes living systems from non-living quite in the same way as Uexküll's watch analogy suggests. The components of a living system are formed during the ontogeny of the individual rather than a priori by an (external) designer. Clocks and basically all artifacts are assemblies of parts that have been built beforehand. In the final product, their organization (i.e. functional relationship) is a result of a designing process. In contrast to that, a living, 'autopoietic' system develops both components and their organizational relation concurrently while growing. Moreover, as the definition of autopoietic systems requires, they never stop producing the components through the interactional network which is formed by these very components. Maturana and Varela speak

[7] See also Susan Oyama (1985) who claims that information is not 'retrieved' but rather 'created' by the system.

[8] This also applies to the development of the apparatus. For example, imprinting (Lorenz, 1982) is nothing else than being temporarily open to a certain kind of perturbation which sustainably and often for the entire life shapes certain parts of the cognitive apparatus. Before and after this period the 'window' is closed.

[9] Of course, this does not necessarily apply to non-cognitive artifacts. A thermostat is a counterexample. It fits the one-dimensional piece of the environment which it senses and effects.

[10] Literally, ''the true is the same as the made'' meaning ''Only what we build we know'' (Glasersfeld, 1995).

of *autopoietic* and *allopoietic* systems, i.e. systems that follow own agendas and systems that are controlled from the outside, respectively.[11] Consequently, current artificial life robots cannot display the characteristics of life, since their (metal) parts are put together similarly as parts of a watch.[12]

How do we have to interpret this sharp distinction of autopoietic vs. allopoietic, and how does the distinction relate to embodiment? Intuitively, any living system is as much a composed system as a watch, although the complexity of the former might be by magnitudes higher. According to the principles of scientific investigation we are not to assume any hidden vitalistic components in living organisms. The components of creatures obey the same physical laws as cogwheels and springs in a watch. Despite the fact that both, natural organisms and artificial aspirants, are 'machines' in the broadest sense, we nevertheless claim a difference. The key is that by letting a system develop on its own, we circumvent problems that typically arise when trying to investigate non-linear systems such as living organisms.

In order to arrive at this interpretation we have to first look at what complexity means. Warren Weaver (1948) provides us with a helpful classification. He distinguishes three kinds of systems. Systems of *simplicity* are championed by classical physics. Typically, we find only a low number of entities (variables) in such systems. Watchmaking and engineering in general, too, work with such systems in which the relationship among their components is linearly decomposable. Simon (1969) has argued that many observable phenomena can be treated like systems of simplicity because they belong to the class of *nearly decomposable* systems in which the interactions among the many subsystems is weak although not negligible. A matrix describing the interactions among the subsystems can be reduced to a sparse matrix, and such a matrix is intellectually and mathematically manageable; no combinatorial explosion arises. The relations amongst the components in a watch belong to this complexity class as much as the computational blockworld Shrdlu.

The second class in Weaver's classification refers to *unorganized complexity*, i.e. systems comprising of myriads of components. Gas is a typical example. While we cannot make statements about individual entities (e.g. gas molecules) we can at least formulate statistical laws regarding the behavior of the entire system (macroscopic properties of gas). Sociology, to some degree, tries to accomplish the same at the level of human populations.

Finally, the case that is interesting for us is called *organized complexity* which appears in systems of a sufficiently high number of variables such that it is not tractable, neither in terms of classical physics nor statistical mechanics. Living cognitive systems belong to this class. It is important to note that we no longer can apply the 'trick' of nearly-decomposing such systems. Too many entities interact with many other entities in non-negligible ways. Ros. Ashby (1973) has an explicitly pessimistic view. He maintains that the scientist who deals with such a complex interactive system '' . . . must be prepared to give up trying to 'understand' it'' (p. 6).

From the perspective of the organized complexity of living systems, it becomes clear that any attempt to design a living/cognitive system is practically impossible. The engineering dichotomy of specifying the components *and* their functional interactions a priori to the actual working of the desired system as a whole can be applied to build watches but is doomed to failure when we want to create living artifacts. The gap between autopoiesis and allopoiesis is not reducible to the actual working of the systems in question. Rather, it results from their historical development in synchronization with their environment.[13]

[11] It may be worth pointing out that Clever Hans, as a living system, was an autopoietic system regardless of whether his behavior seemed being controlled from the outside. Recognizing the right cues enabled him to behave in a way which ultimately got him a reward. He was embodied in the physical space (as any other autopoietic system according to the definition given by Maturana and Varela (1980)). As already stated before, he was not embodied in the intellectual domain of mathematics. In this regard it also doesn't make sense to speak in terms of goal-directed behavior since horses do not deliberately and consciously set goals.

[12] Some authors, e.g. Margaret Boden (1999), argue that if we want to create artificial life, we must, therefore, build a biochemical one.

[13] Note that autopoiesis does not mean the same as living system (however, every living organism is an autopoietic system). And while there is no reason to assume that artifacts cannot undergo historical developments, this ability is even crucial for cognitive artifacts.

How does this relate to embodiment? Clearly, the Quick et al. minimal definition of embodiment does not take the historicity of living organisms into consideration. Every system, autopoietic or allopoietic, is embodied in the sense that it is subject to perturbations from an outside (think of Rembrandt's fly). Embodiment of autopoietic systems does not only mean that the components and their organizational network has been shaped through the interaction with their environment. It also brought forward the aspect of self-steering in the sense of autopoiesis. In other words, the embodiment of a system is synonymous with competence in its environment. Brooks made a first albeit not sufficiently big step in this direction. He recognized that if we want robots to be able to navigate in a physical environment it is necessary to let them deal with ''the here and now of the world directly influencing the behavior of the system'' (Brooks, 1991, p. 571) rather than with abstract descriptions of the world. In this situation, robots ''experience the world directly—their actions are part of a dynamic with the world and have immediate feedback on their own sensations''.[14]

From what has been said above it is obvious that this form of Brooksian embodiment[15] treats the agent as being at the mercy of its environment. The agent is embedded/situated in the stream of environmental events. Although Brooks implements goals (such as coda-can collecting) into his robots, they are typically implemented causally rather than explicitly ('physical gounding', Brooks, 1993). This definition simply equates embodiment with situatedness (or embeddedness) in complex environments. Critics such as Etzioni are right that of course a UNIX environment too fulfils the criteria of this kind of embodiment. It provides enough complexity in order to go beyond what a logical–formalistic approach can cope with. And it separates the designer of the agent from the designer of the environment, thus providing enough 'surprises' in addition. However, such a UNIX-agent can no more transcend the problem of being designed (in contrast to being co-evolved) than a Brooksian robot.

In both, Brooks and Etzioni agents, the usefulness of artifacts is emphasized, be it as can-collecting agent or as help and surveillance system in a computational environment. But how do task, purpose, and goal indeed relate to embodiment?

7. Goals in embodied systems

Rolf Pfeifer and Christien Scheier (1999) write that one goal of embodied cognitive science is ''building an agent for a particular task'' (Pfeifer and Scheier, p. 649). As humans, we don't have problems to impose our goals on other entities, whether horses, airplanes, or mathematical formulae. These entities are tamed to carry out our orders. Some might even say that mankind tamed Nature in general. The usage of mathematical equations demonstrates this habit quite clearly without invoking any 'anthropocentric distortions'. We (usually) know what the meaning of the labels (variables) is even though we might not be the author of the equations. For example, calculating $F = gm_1m_2/r^2$, we know that providing numerical values for the variables on the right side (input) will produce a value for F (the output) at the other side. The equation has been designed to serve our purposes. It would be ridiculous to say that the equation had a goal of its own. Nor would it be appropriate to say that it emerged in a process of self-organization. Physicists deliberately wrote it down in order to express a relationship in their experiences when interacting with the world, namely the relationship between masses, distance and force between two physical objects. In this perspective, mathematical formulae are purposeful agents embedded—rather than embodied—in the environment of mathematics.

And what is more important, we completely understand the working of the formula. As I said, a mathematical formula doesn't provide any grip for

[14] The most extreme position is to substitute cognitive capabilities by exploiting the physical structure of robots and self-organizing properties of group processes. For example, René te Boekhorst and Marinus Maris describe a collective heap building process by a group of robots (Maris & te Boekhorst, 1996). The sensory input of the robots was restricted such that they could only detect obstacles which are diagonally to the left or right but not in front of the robot. In this way, the robots collide with objects that are exactly in front of them, thus pushing them and forming clusters.

[15] Or 'Loebian/mechanistic embodiment' in Sharkey and Ziemke's (2000) terminology.

anthropomorphic distortion. Things are becoming slightly less clear when looking at physical instantiations of simple functions. The often cited thermostat is one example where—trivially speaking—we could already start to ascribe some inherent goals to this simple machine. In colloquial speech, the more complex a system becomes, the more it hides its functioning and internal mechanisms from the curious observer, the more likely we ascribe purposeful behavior to it. In man-made (i.e. allopoietic) machines designed so far, the purpose lies exclusively in the domain of descriptions of the observer. The artifact itself has as few goals as a mathematical formula.

However, this aspect is easily overlooked. As argued in Riegler (1997) and Peschl and Riegler (1999), a typical deficiency of many artificial life models is the *PacMan syndrome* which results from the 'designedness' of the system. Artificial agents interact with anthropomorphically defined entities, such as 'food' and 'enemy', which make sense only to the programmer of the system. Such models perform a mere optimizing task which yields a maximum gain of energy together with a minimum loss of health. No attention is paid to questions like: how have organisms arrived at the idea that something is a source of food? How do they 'know' that another creature is a dangerous opponent? Predators do not have signs on their backs saying 'I'm your enemy'. Consequently, the part of the computer program which implements the agent has no 'internal drive', no goal to avoid 'enemies' and to approach 'food'.

From an engineering point of view it is not clear how these goals can be defined a priori. We have seen that for cognitive systems their goals have to build up within their development (including phylogeny). Since internal goals are inherently interlinked with embodiment, the design of purpose is as impossible as the design of an embodied agent. If we want embodied artificial agents to do something useful, the only possible way is to do the same as humans have been doing when they tamed wild animals. That is, to engage in structural coupling with them in order to establish, what Maturana and Varela (1980) call, a *consensual domain*. Participants in a consensual domain become involved in an ongoing mutual process of triggering compensating perturbations in one another. This process enabled the behavior of animals to be shaped, and it may be an appropriate way to do the same with otherwise embodied autonomous cognitive artifacts.

8. Conclusions

A system is embodied if it has gained competence within the environment in which it has developed. In the case of the physical domain, living organisms are embodied. Embodiment requires structural coupling between system and environment, i.e. the system must be able to engage in a mutual sequence of perturbations with its environment. This definition of embodiment does not only apply to the physical domain. Computer programs may also become embodied. The fact that most or all current artificial intelligence programs do not exhibit embodiment has to do with their explicit design rather than with the space they are habitating. Designed systems necessarily always include the goal of the designer as their main driving instance. Such artifacts are built as purposeful systems since the specification requires the dualism of a priori defining the components and their interactional relationship *before* the entire system starts to work. In this sense they are not embodied, not 'synchronized' with but merely embedded in the dynamics of their environment. From this view, embodiment becomes a matter of complexity. It does not introduce a new quality into cognitive science. Rather, it reflects the difficulties of a human designer to cope with organized complexity. Due to these limitations any designing process of systems of non-trivial complexity must remain incomplete, thus preventing the system from gaining cognitive autonomy. The lesson embodiment teaches us should be taken seriously; however there is no reason to limit the creation of cognitive artifacts only to physical instantiations such as robots.

Acknowledgements

I acknowledge financial support by the Flamish Fonds for Scientific Research FWO. I also would like to thank the anonymous reviewers for their comments.

References

Ashby, W. R. (1973). Some peculiarities of complex systems. *Cybernetic Medicine*, *9*, 1–7.

Boden, M. A. (1999). Is metabolism necessary? *British Journal for the Philosophy of Science*, *50*(2), 231–248.

Brooks, R. A. (1991). Intelligence without reason. In *Proceedings of the twelfth international joint conference on artificial intelligence*. San Mateo, CA: Morgan Kaufmann, pp. 569–595.

Brooks, R. A. (1993). The engineering of physical grounding. In *Proceedings of the fifteenth annual conference of the cognitive science society*. Hillsdale, NJ: Lawrence Erlbaum, pp. 153–154.

Dautenhahn, K. (1997). I could be you—the phenomenological dimension of social understanding. *Cybernetics and Systems*, *25*(8), 417–453.

Dennett, D. C. (1984). Cognitive wheels: the frame problem of AI. In Hookway, C. (Ed.), *Minds, machines, and evolution: philosophical studies*. London: Cambridge University Press, pp. 129–151.

Etzioni, O. (1993). Intelligence without robots: a reply to Brooks. *AI Magazine*, *14*(4), 7–13.

Feynman, R. (1985). *Surely you're joking, Mr. Feynman!*. New York: W.W. Norton.

Franklin, S. (1995). *Artificial minds*. Cambridge, MA: MIT Press.

Franklin, S. (1997). Autonomous agents as embodied AI. *Cybernetics and Systems*, *28*(6), 499–520.

Glasersfeld, E. von (1983). Learning as a constructive activity. In Bergeron, J. C., & Herscovics, N. (Eds.), *Proceedings of the fifth annual meeting of the North American chapter of the international group for the psychology of mathematics education*. Montreal: University of Montreal, pp. 41–69.

Glasersfeld, E. von (1995). *Radical constructivism. A way of knowing and learning*. London: Falmer Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.

Heschl, A. (1990). L = C. A simple equation with astonishing consequences. *Journal of Theoretical Biology*, *145*, 13–40.

Lorenz, K. (1973). *Die Rückseite des Spiegels. Versuch einer Naturgeschichte menschlichen Erkennens*. Piper, München. English translation: Lorenz, K. (1976). *Behind the mirror: a search for a natural history of human knowledge*. Ronald Taylor, trans. New York: Harcourt Brace Jovanovich.

Lorenz, K. (1982). Kant's doctrine of the a priori in the light of contemporary biology. In H. C. Plotkin (Ed.), *Learning, development and culture* (pp. 121–143). Chichester: John Wiley. German original: Lorenz, K. (1941). *Kant's Lehre vom Apriorischen im Lichte gegenwärtiger Biologie*. Blätter für Deutsche Philosophie 15, pp. 94–125.

Maris, M., & te Boekhorst, R. (1996). Exploiting physical constraints: heap formation through behavioral error in a group of robots. In Asada, M. (Ed.), *Proceedings of IROS'96: IEEE/RSJ international conference on intelligent robots and systems*, pp. 1655–1660.

Maturana, H. R. (1980). Autopoiesis: reproduction, heredity and evolution. In Zeleny, M. (Ed.), *Autopoiesis, dissipative structures and spontaneous social orders*, *AAAS selected symposium 55*. Boulder, CO: Westview.

Maturana, H. R., & Varela, F. J. (1980). In *Autopoiesis and cognition: the realization of the living*, *Boston studies in the philosophy of sciences*, vol. 42. Boston: D. Reidel.

Neisser, U. (1976). *Cognition and reality*. San Francisco: W.H. Freeman.

Oyama, S. (1985). *The ontogeny of information: developmental systems and evolution*. Cambridge, Cambridge University Press.

Peschl, M., & Riegler, A. (1999). Does representation need reality? In Riegler, A., Peschl, M., & von Stein, A. (Eds.), *Understanding representation in the cognitive sciences*. New York: Kluwer Academic/Plenum, pp. 9–17.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge, MA: MIT Press.

Piaget, J. (1954). *The construction of reality in the child*. New York: Ballentine.

Popper, K. R. (1979). In 5th rev. ed, *Objective knowledge: an evolutionary approach*. Oxford: Clarendon Press.

Quick, T., Dautenhahn, K., Nehaniv, C., & Roberts, G. (1999). The essence of embodiment: a framework for understanding and exploiting structural coupling between system and environment. In Dubois, D. M. (Ed.), *Proceedings of the 3rd international conference on computing anticipatory systems (CASYS'99)*, *AIP conference proceedings*, vol. 517. Heidelberg: Springer Verlag.

Riegler, A. (1994). Constructivist artificial life: the constructivist–anticipatory principle and functional coupling. In Hopf, J. (Ed.), *Genetic algorithms within the framework of evolutionary computation*. Saarbrücken: Max-Planck-Institut für Informatik, pp. 73–83, MPI-I-94-241.

Riegler, A. (1997). Ein kybernetisch-konstruktivistisches Modell der Kognition. In Müller, A., Müller, K. H., & Stadler, F. (Eds.), *Konstruktivismus und Kognitionswissenschaft. Kulturelle Wurzeln und Ergebnisse*. Wien: Springer, pp. 75–88.

Riegler, A. (1998). The end of science: can we overcome cognitive limitations? *Evolution & Cognition*, *4*(1), 37–50.

Riegler, A. (2001a). Towards a radical constructivist understanding of science. *Foundations of Science*, *6*(1), 1–30.

Riegler, A. (2001b). The role of anticipation in cognition. In Dubois, D. M. (Ed.), *Computing anticipatory systems*, *Proceedings of the American institute of physics*, vol. 573, pp. 534–541.

Sharkey, N., & Ziemke, T. (2000). Life, mind and robots – the ins and outs of embodied cognition. In Wermter, S., & Sun, R. (Eds.), *Hybrid neural systems*. Heidelberg: Springer Verlag.

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge: MIT Press.

Stewart, J. (1996). Cognition = life: implications for higher-level cognition. *Behavioural Processes*, *35*, 311–326.

Uexküll, J. von (1934). *Streifzüge durch die Umwelten von Tieren und Menschen: Ein Bilderbuch unsichtbarer Welten*. Berlin: J. Springer. Engl. translation: Uexküll, J. von (1957). A stroll through the worlds of animals and men. In: Schiller, C. H. (ed. and transl.) *Instinctive behavior: the development of a modern concept* (pp. 5–80). New York: International Universities Press.

Weaver, W. (1948). Science and complexity. *American Scientist*, *36*, 536–544.

Winograd, T. (1972). *Understanding natural language*. New York: Academic Press.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition*. Norwood: Ablex.