Are Robots Embodied?

Tom Ziemke
Department of Computer Science, University of Skövde, Sweden, tom@ida.his.se

Abstract

Embodiment has become an important concept in many areas of cognitive science. There are, however, very different notions of exactly what embodiment is and what kind of body is required for what kind of embodied cognition. Hence, while many would agree that humans are embodied cognizers, there is much less agreement on what kind of artefact could be considered as embodied. This paper identifies and contrasts five different notions of embodiment which can roughly be characterized as (1) *structural coupling* between agent and environment, (2) *historical embodiment* as the result of a history of structural coupling, (3) *physical embodiment*, (4) *'organismoid' embodiment*, i.e. organism-like bodily form (e.g., humanoid robots), and (5) *organismic embodiment* of autopoietic, living systems.

1. Introduction

The concept of embodiment has since the mid-1980s been used extensively in the cognitive science and AI literature, in such terms as 'Embodied Mind' (e.g. Varela *et al.*, 1991; Lakoff & Johnson, 1999), 'Embodied Intelligence' (e.g. Brooks, 1991), 'Embodied Action' (Varela *et al.*, 1991), 'Embodied Cognition' (e.g. Clark, 1997), 'Embodied AI' (e.g. Franklin, 1997), and 'Embodied Cognitive Science' (Pfeifer & Scheier, 1999; Clark, 1999). Furthermore, there obviously are different types and notions of embodiment as can be seen in the variety of terms such as 'situated embodiment' (Zlatev, 1997), 'mechanistic embodiment' (Sharkey & Ziemke, 2000, in press), 'phenomenal embodiment' (Sharkey & Ziemke, 2000, in press), 'natural embodiment' (Ziemke, 1999), 'naturalistic embodiment' (Zlatev, 2001), 'social embodiment' (e.g. Duffy, 2001), plus in this paper 'physical embodiment', 'organismoid embodiment', and 'organismic embodiment'.

Embodiment is nowadays by many researchers considered a *conditio sine qua non* for any form of natural or artificial intelligence. Pfeifer and Scheier (1999), for example, argued:

> *[I}ntelligence cannot merely exist in the form of an abstract algorithm but requires a physical instantiation, a body.*

Embodiment is commonly considered one of the key ideas that distinguishes recent work on situated/embodied/embedded/interactive theories of cognition from the approach of classical cognitive science which, based on functionalism, had its focus on 'disembodied' computation (cf., e.g., Varela *et al.*, 1991; Clark, 1997; Pfeifer & Scheier, 1999). However, while many researchers nowadays agree that cognition has to be embodied, it is less clear so far what exactly that means. Wilson (submitted) has recently pointed out that the diversity of claims in the field is problematic:

> *While this general approach [of embodied cognition or embodied cognitive science] is enjoying increasingly broad support, there is in fact a great deal of diversity in the claims involved and the degree of controversy they attract. If the term "embodied cognition" is to retain meaningful use, we need to disentangle and evaluate these diverse claims.*

In particular it is actually far from clear what kind of body is required for embodied cognition. Hence, while it might be agreed upon that humans are embodied cognizers, there is little agreement on what kind of body an artificial intelligence would have to be equipped with. As discussed in detail in Section 4, Pfeifer and Scheier's (1999) above view, for example, that intelligence requires a *physical* body, is not at all generally accepted.

This paper aims to identify and discuss a number of different notions of embodiment in the cognitive science and AI literature. Due to space restrictions, none of these notions is here argued for or against in particularly much detail, although admittedly the last and most restrictive notion, which we refer to as 'organismic embodiment', does receive much more attention than the others, since it is derived from our own earlier work (Sharkey & Ziemke, 1998, 2000, in press; Ziemke, 1999, 2000, 2001a; Ziemke & Sharkey, 2001). The rest of this paper is structured as follows: The next section briefly recapitulates Searle's (1980) Chinese Room Argument as it is an important part of the background of current work in embodied cognitive science and will help us to assess the state of the art later in this paper. Section 3 then briefly discusses different views of embodied cognition, following the distinctions made by Wilson (submitted). Section 4 identifies five different notions of embodiment and exactly what kind of body is required for embodied cognition. Section 5, finally, presents a brief summary.

[*] This paper is a revised version of Ziemke (2001b). Parts of Section 4.5 have also appeared in Ziemke (2001a) and/or Ziemke & Sharkey (2001).

2. Background: The Chinese Room

An important part of the scientific background of today's work on embodied cognition in general, and adaptive/cognitive/epigenetic robotics in particular, is Searle's (1980) *Chinese Room Argument* (CRA). In his critique of predominantly computationalist AI research at the time Searle distinguished the positions of

- *weak* or *cautious AI* which sees the computer as a powerful tool in the study of mind (a position he agreed with), and
- *strong AI* which would hold that "the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states".

Searle's now famous CRA against the position of strong AI goes approximately like this: Suppose you, knowing no Chinese at all, are locked in a room. Under the door you are passed some Chinese writing. With you in the room you have some instructions (in English) which allow you to manipulate and return Chinese symbols in a purely syntactic fashion. Now, the crucial question is, do you *understand* Chinese in the sense that you actually know what any of the symbols mean? The obvious answer, Searle argued, is that you do not. However, Chinese-speaking observers outside the room might consider the symbols sent into the room questions and, assuming the English instructions are very good, the returned symbols correct answers. Searle's point was that, although from outside the room you might be considered to understand, obviously everybody who knows what goes on inside the room realizes that you are just "manipulating uninterpreted formal symbols". Furthermore, Searle concluded, since you, inside the room, are "simply an instantiation of the computer program", any computer using a purely formally defined program would have to be said to understand as much of what it processes as you understand Chinese; namely nothing at all.

The reason for this lack of understanding in the computer's case, Searle elaborated, is that, due to the fact that there are no causal connections between the internal symbols and the external world they are supposed to represent, purely computational AI systems lack *intentionality*. In other words, they do not have the capacity to relate their internal processes and representations to the external world. Searle argued that the main reason for the failure of strong (traditional) AI was that it is concerned with *computer programs*, but "has nothing to tell us about *machines*" (Searle, 1980), i.e. physical systems situated in and causally connected to their environments. In fact, he accused AI of dualism, for its belief that disembodied, i.e. body-less and body-independent, computer programs could be intelligent. Hence, his conclusion was that AI research, instead of focusing on purely formally defined computer programs, should be working with physical machines equipped with (some of) the causal powers of living brains/organisms, including perception, action and learning. In fact, that is approximately what modern AI, in particular work in adaptive/cognitive/epigenetic robotics, does. It focuses on robots, i.e. (physically) embodied systems, which 'perceive', 'act' and 'learn' (by artificial means) in interaction with the environment they are situated in, and thus might be argued to develop their *own*, *observer-independent* representations and intentionality. We will return to the question whether or not this argument holds in the end of this paper.

3. Different Views of Embodied Cognition

Wilson (submitted) recently distinguished between six different views of embodied cognition, of which, however, only one explicitly addresses the role of body:

- "Cognition is situated": This claim is obviously widely held in the literature on embodied cognition[1]. Wilson herself distinguished between situated cognition, which takes place "in the context of task-relevant inputs and outputs", and "off-line cognition", which does not.
- "Cognition is time-pressured": That means, cognition is constrained by the requirements of real-time interaction with the environment, e.g. the 'representational bottleneck' (e.g. Brooks, 1991; Clark, 1997; Pfeifer & Scheier, 1999).
- "We off-load cognitive work onto the environment": Brooks (1991) formulated a similar claim by saying that "the world is its own best model". A well-known example is Kirsh & Maglio's (1994) study of 'epistemic actions' in the game of Tetris, i.e. decision-preparing movements carried out in the world, rather than in the head.
- "The environment is part of the cognitive system": An example of this view could be Hutchins' (1995) work on distributed cognition, in which, for example, the instruments in a cockpit are considered parts of the cognitive system. However, as Wilson points out, "relatively few theorists appear to hold consistently to this position in its strong form".
- "Cognition is for action": A claim made, for example, by Franklin (1995), who argued that minds are the control structures of autonomous agents.
- "Off-line cognition is body-based": According to Wilson, this claim has so far received least attention in the cognitive science literature, although "it may in fact be the best documented and most powerful of the six claims". Perhaps the most prominent example is the work of Lakoff & Johnson (1980, 1999) who have

[1] It might be worth pointing out though that the concept of situatedness itself is far from being well defined (cf., e.g., Ziemke, 2000, 2001a).

argued that abstract concepts are based on metaphors grounded in bodily experience/activity. This claim is discussed in further detail in the following section.

4. Different Notions of Embodiment

As noted in the previous section, perhaps somewhat surprisingly, many discussions/notions of em*bodi*ed cognition actually pay relatively little attention to the nature and the role of the body involved (if at all). Only Wilson's sixth view, of 'off-line cognition' as body-based, explicitly mentions the body as playing a central role. It does, however, leave open the question whether, for example, a humanoid robot, i.e. a robot with more or less roughly human-like form, could have the same type of cognition as its living counterpart.

We here would like to distinguish between the following five notions of what kind of body/embodiment is required for (embodied) cognition:

- *structural coupling* between agent and environment,
- *historical embodiment* as the result of a history of structural coupling
- *physical embodiment*,
- *'organismoid' embodiment*, i.e. organism-like bodily form (e.g., humanoid robots), and
- *organismic embodiment* of autopoietic, living systems

Each of the above notions of embodiment is in the following elaborated in a separate subsection. It might be worth pointing out beforehand that at least some of them are actually groups of more or less closely related notions rather than single, well-defined positions.

4.1. *Embodiment as 'Structural Coupling'*

The probably broadest notion of embodiment is that systems are embodied if they are 'structurally coupled' to their environment. Note that this does not necessarily require a body. Franklin (1997), for example, argued:

> *Software systems with no body in the usual physical sense can be intelligent. But they must be embodied in the situated sense of being autonomous agents structurally coupled with their environment.*

The concept of *structural coupling* originates from Maturana and Varela's (1980, 1987) work on the biology of cognition, which will be discussed in further detail in Section 4.5. Inspired by this concept, Quick & Dautenhahn (1999)[2] have attempted to provide a "precise definition" of embodiment:

> *A system X is embodied in an environment E if perturbatory channels exist between the two. That means, X is embodied in E if for every time t at which both X and E exist, some subset of E's possible states with respect to X have the capacity to perturb X's state, and some subset of X's possible states with respect to E have the capacity to perturb E's state.*

It could be argued that this definition, which Quick & Dautenhahn referred to as "minimal", is of limited use to cognitive science due to the fact that it is not particularly restrictive. Riegler (in press), for example, argued that "[w]hile this attempt to clarify the notion of embodiment is an important first step, it is at the same time an insufficient characterization" due to the fact that "every system is in one sense or another structurally coupled with its environment". That means, this definition of embodiment does not make a distinction between cognitive and non-cognitive systems, which can be illustrated with Quick & Dautenhahn's (1999) example of an granite outcrop (X) on the Antarctic tundra (E). The outcrop is persistently perturbed by the wind, and in turn perturbs the air-currents' flow. Hence, it is an embodied system according to the above definition, although certainly not many cognitive scientists would actually consider this an example of embodied cognition.

4.2. *Historical Embodiment*

Several researchers have emphasized that cognitive systems are not only structurally coupled to their environment in the present, but that their embodiment is in fact a result or reflection of a *history* of agent-environment interaction. Varela *et al.* (1991), for example, argued:

> *[K]nowledge depends on being in a world that is inseparable from our bodies, our language, and our social history – in short, from our embodiment.*

Ziemke (1999) pointed out :

> *Natural embodiment [of living systems]… reflects/embodies the history of structural coupling and mutual specification between agent and environment in the course of which the body has been constructed.*

Similarly, Riegler (in press) includes the agent's adaptation to its environment in his definition of embodiment: "A system is embodied if it has gained competence within the environment in which it has developed".

4.3. *Physical Embodiment*

A notion of embodiment, which excludes the software agents Franklin (1997) considered as embodied (cf. Section 4.1), is the view that embodied systems need to have a "physical instantiation" in the sense of Pfeifer and Scheier (cf. Section 1), i.e. a physical body. Again,

[2] See also Quick *et al.* (1999).

in itself, this is not particularly restrictive and it still includes Quick & Dautenhahn's (1999) above granite outcrop.

It might be worth pointing out that although historical embodiment and physical embodiment both can be considered special cases of structural coupling, none of these two notions includes or excludes the other. Riegler (in press), for example, noted that his definition of embodiment (cf. Section 4.2) "does not exclude domains other than the physical domain"; in particular "[c]omputer programs may also become embodied" if they are the result of self-organization rather than explicit design. Similarly, living systems are examples of physically embodied systems which are also 'historically embodied', whereas many other physical systems are not.

A somewhat more restrictive version of the notion of physical embodiment[3] is the view that embodied systems should be connected to their environment not just through physical forces, but also through sensors and motors. From an AI perspective, Brooks (1990), for example, formulated the 'Physical Grounding Hypothesis':

> *Nouvelle AI is based on the physical grounding hypothesis. This hypothesis states that to build a system that is intelligent it is necessary to have its representations grounded in the physical world. ... To build a system based on the physical grounding hypothesis it is necessary to connect it to the world via a set of sensors and actuators.*

4.4. 'Organismoid' Embodiment

Another, yet more restrictive notion of physical embodiment is that at least certain types of organism-like cognition might be limited to organism-like bodies, i.e. physical bodies which at least to some degree have the same or similar form and sensorimotor capacities as living bodies. It should be noted that the notion of 'organismoid' embodiment here is intended to cover both living organisms and their artificial counterparts. [4]

One of the simplest examples of organism-like embodiment might be the Khepera robot used by Lund *et al.* (1998). It was equipped with an additional auditory circuit and two microphones which had the same distance from each other as the two 'ears' of the crickets whose phonotaxis it was supposed to model. In this case the placement of the sensors, in both cricket and robot, reduced the amount of internal processing required to respond selectively to certain sound frequencies. Note that in this case the bodies of the cricket and the wheeled robot were in fact *very* different, except for one crucial detail, the distance between the 'ears'.

The most prominent, and perhaps the most complex, example of artificial organismoid embodiment are humanoid robots such as the famous Cog (Brooks & Stein, 1993; Brooks *et al.*, 1998), based on the argument that research in AI and cognitive robotics, in order to be able to address/investigate human-level cognition, has to deal with human-like artefacts. [5]

Dreyfus (1996), for example, pointed out that "there are many important ways in which neural nets differ from embodied brains". He argued that neural nets would need to be "put into [humanoid] robots" since the lack of body and environment

> *… puts disembodied neural-networks at a serious disadvantage when it comes to learning to cope in the human world. Nothing is more alien to our life-form than a network with no up/down, front/back orientation, no interior/exterior distinction, … The odds against such a net being able to generalize as we do, … are overwhelming.*

This argument is closely related to Wilson's sixth view of embodied cognition (cf. previous section) and, for example, the aforementioned work of Lakoff & Johnson (1980, 1999) on the bodily/metaphorical basis of abstract concepts. Lakoff (1988) summarized the basic idea as follows:

> *Meaningful conceptual structures arise from two sources: (1) from the structured nature of bodily and social experience and (2) from our innate capacity to imaginatively project from certain well-structured aspects of bodily and interactional experience to abstract conceptual structures.*

If, for example, the concept of 'grasping an idea' is grounded in the bodily experience/activity of grasping physical objects, then a robot without any gripper arm/hand could hardly be expected to be able to understand that concept. A similar argument has been presented by Keijzer (1998) who has questioned the suitability of wheeled robots for the study of the behavior/cognition of organisms with completely different means of locomotion.

---

[3] This could be considered an independent notion, perhaps with the label *'sensorimotor embodiment'*. However, since it seems rather difficult to define exactly what 'sensors' and 'motors' are and how they differ from other 'perturbatory channels', we abstain from doing so in this paper.

[4] It might be worth pointing out that while living and artificial 'organismoids' can be considered physically embodied, only the former are necessarily 'historically embodied' (cf. Section 4.2) whereas the latter are usually results of human design, although typically biologically inspired.

[5] Hence, 'humanoid embodiment' could be considered a special case of 'organismoid embodiment', which might be of particular interest to cognitive science. It should be noted, however, that this leaves open the question what exactly the supposedly cognition-relevant bodily differences between humans and other animals/organisms are.

4.5. Organismic Embodiment

The most restrictive notion of embodiment discussed in this paper holds that cognition is not only limited to bodies of organism-like form, but in fact to organisms, i.e. *living bodies*. This notion has its roots in the work of theoretical biologist von Uexküll (1928, 1982) and its modern counterpart, the work of Maturana & Varela (1980, 1987) on the biology of cognition, which holds, roughly speaking, that cognition is what living systems do in interaction with their environment.[6] According to this view, there is a clear difference between living organisms, which are *autonomous* and *autopoietic*, and man-made machines, which are heteronomous and allopoietic (cf. Sharkey & Ziemke, 2000, in press; Ziemke & Sharkey, 2001).

Von Uexküll (1928), for example, argued that all action of organisms is a mapping between individual stimuli and effects, depending on an historically created basis of reaction (*Reaktionsbasis*), i.e. a context-dependent behavioral disposition. Machines, on the other hand, at least in von Uexküll's time (1864-1944), did not have such an historical basis of reaction, which, according to von Uexküll, can only be grown - and there is no growth in machines. Von Uexküll further elaborated that the rules machines follow are not capable of change, due to the fact that machines are fixed structures. That means, the rules that guide their operation, are not their 'own' but human rules, which have been built into the machine, and therefore also can be changed only by humans, i.e. mechanisms are *heteronomous*. Machines could therefore, according to von Uexküll, when they get damaged, not repair or regenerate themselves. Living organisms, on the other hand, can, because they contain their functional rule (*Funktionsregel*) themselves, and they have the protoplasmic material, which the functional rule can use to fix the damage autonomously. This can be summarized by saying that *machines act according to plans* (their human designers'), whereas *living organisms are acting plans* (von Uexküll, 1928).

This is also closely related to what von Uexküll (1982) called the "principal difference between the construction of a mechanism and a living organism":

> *Every machine, a pocket watch for example, is always constructed centripetally. In other words, the individual parts of the watch, such as its hands, springs, wheels, and cogs, must always be produced first, so that they may be added to a common centerpiece.*
>
> *In contrast, the construction of an animal, for example, a triton, always starts centrifugally from a single cell, which first develops into a gastrula, and then into more and more new organ buds.*
>
> *In both cases, the transformation underlies a plan: the 'watch-plan' proceeds centripetally and the 'triton-plan' centrifugally. Two completely opposite principles govern the joining of the parts of the two objects.*

Maturana and Varela (1980, 1987) distinguished between the organization of a system and its structure. The *organization*, similar to von Uexküll's notion of a building-plan (*Bauplan*), denotes "those relations that must exist among the components of a system for it to be a member of a specific class" (Maturana and Varela, 1987). Living systems are characterized by their autopoietic organization. An *autopoietic* system is a special type of homeostatic machine for which the fundamental variable to be maintained constant is its own organization (unlike regular homeostatic machines, which typically maintain single variables, such as temperature or pressure). A system's *structure*, on the other hand, denotes "the components and relations that actually constitute a particular unity, and make its organization real" (Maturana and Varela, 1987). Thus the structure of an autopoietic system is the concrete realization of the actual components and the actual relations between them. Its organization is constituted by the relations between the components that define it as a unity of a particular kind. These relations are a network of processes of production that, through transformation and destruction, produce the components themselves. It is the interactions and transformations of the components that continuously regenerate and realize the network of processes that produced them.

Hence, according to Maturana and Varela (1980), living systems are not at all the same as machines made by humans, such as cars and robots, which are *allopoietic*. Unlike an autopoietic machine, the organization of an allopoietic machine is given in terms of a concatenation of processes. These processes are not the processes of production of the components that specify the machine as a unity. Instead, its components are produced by other processes that are independent of the organization of the machine. Thus the changes that an allopoietic machine goes through without losing its defining organization are necessarily subordinated to the production of something different from itself. In other words, it is not truly autonomous, but heteronomous. In contrast, a living system is truly autonomous in the sense that it is an autopoietic machine whose function it is to create and maintain the unity that distinguishes it from the medium in which it exists. It is worth pointing out that, despite differences in terminology, Maturana and Varela's distinction between autopoietic and allopoietic machines, is very similar to von Uexküll's (1928) distinction between human-made mechanisms, which are constructed centripetally by a designer and act according to his/her plan, and organisms, which as 'living plans' 'construct' themselves in a centrifugal fashion.

[6] See also Lemmen (1996) and Stewart (1996), who summarizes this view as "Cognition = Life".

As discussed in detail elsewhere (Ziemke, 2000, 2001a; Ziemke & Sharkey, 2001), much progress has been made in the direction of self-organizing robots in recent AI and artificial life research. Unlike the machines in von Uexküll's time, today's adaptive robots can 'grow' in interaction with their environment through the use of artificial evolutionary and learning techniques. Furthermore, robot bodies can be evolved centrifugally over generations in some sense (e.g. Lipson & Pollack, 2000).

Nevertheless, in all current experimental work the robot's embodiment and its relation to its environment are very different from the living organism's. A robot might have adapted its control system, possibly even its physical structure to some degree, in interaction with its environment, and thus might be argued to have acquired a certain degree of "epistemic autonomy" (Prem, 1997; cf. also Cariani, 1992). This (partial) self-organization, however, typically starts and ends with a bunch of physical parts and a computer program. Furthermore, the process is determined, started and evaluated by a human designer, i.e. the drive to self-organize does not actually lie in the robot's components themselves and success or failure of the process is not 'judged' by them either. For example, in the above case of the evolution of robot bodies (e.g. Lipson & Pollack, 2000), there is no growth or adaptation of the individual robot body. Instead body plans are first evolved in the computer, i.e. 'outside' the robot, and then implemented in a robot body. In von Uexküll's terms, the evolution of the body plan might have followed centrifugal principles, the resulting robot bodies are, however, still built in a centripetal fashion and from then on can no longer self-organize. Hence, these bodies are not at all 'living plans' in von Uexküll's sense, which construct themselves, but they still are constructed according to an extrinsic plan. The components might be better integrated after having self-organized; they might even be considered 'more autonomous' for that reason, but they certainly do not become alive in that process, i.e. they remain allopoietic rather than autopoietic.

Any living organism, on the other hand, starts its self-organizing process from a single autonomous cellular unity (*Zellautonom*). The drive to self-organize is part of its 'building plan' (*Bauplan*), and it is equipped, in itself, with the resources to 'carry out that plan'. From the very beginning the organism is a viable unity, and it will remain that throughout the self-organizing process (until it dies). Today's adaptive robots, however, although self-organizing in some sense, using artificial evolutionary and learning techniques, have no intrinsic needs that the self-organizing process would have to fulfill to remain 'viable'. Following the arguments of von Uexküll as well as Maturana and Varela, every organism can be considered a living, self-constructing and self-modifying 'hypothesis' in von Uexküll's sense of an 'acting plan', maintaining its viability through adaptation under environmental constraints (cf. Ziemke, 2001). Hence, in the organism's case viability in the biological sense of survival and viability in the sense of fit between behavioral/conceptual mechanisms and experience are closely connected. A robot, on the other hand, 'lacks' the *intrinsic* requirement of biological viability. Hence, the viability of its behavioral/conceptual mechanisms can ultimately always only be evaluated from the outside (with respect to fitness function, reinforcement, error measures, etc.). Thus, for the robot the only criterion of success or failure is still the designer's and/or observer's evaluation or interpretation, i.e. this criterion is entirely *extrinsic* to the robot.

A key problem with much current research in cognitive/adaptive/epigenetic robotics, we believe, is that, despite claims to the contrary and despite the emphasis of 'embodiment', many researchers are still devoted to the computationalist/functionalist hardware-software distinction. Much research effort is spent on adaptive control mechanisms and how to achieve certain behaviors in robots through self-organization of these control mechanisms, reducing the body to the computational control system's sensorimotor interface to the environment. Maturana and Varela (1987), however, have argued, that living bodies and nervous systems are not at all separate parts:

> *... the nervous system contains millions of cells, but all are integrated as components of the organism. Losing sight of the organic roots of the nervous system is one of the major sources of confusion when we try to understand its effective operation.*

Similarly, T. von Uexküll *et al.* (1993), in their discussion of *endosemiosis* (sign processes inside the organism), point out that the living body, which we experience to be the center of our subjective reality (*Wirklichkeit*), is the correlate of a neural *counterbody* (*Gegenkörper*) which is formed and updated in our brain as a result of the continual information flow of proprioceptive signs from the muscles, joints and other parts of our limbs. This neural counterbody is the center of the so-called *neural counterworld* (cf. von Uexküll, 1909, 1985), created and adapted by the brain from the continual stream of signs from the sensory organs. According to T. von Uexküll *et al.*, counterbody and counterworld form an undividable unity, due to the fact that all processes/events we perceive in the world really are 'countereffects' to real or potential effects of our motor-system, and together with these they form the spatial structure within which we orient ourselves. A robot, on the other hand, has no endosemiosis whatsoever in the body (its physical components) as such. Thus, there is no integration, communication or mutual influence of any kind between parts of the body, except for their purely mechanical interaction. Further, there is no meaningful integration of the 'artificial nervous system' and the physical body, beyond the fact that some parts of the body provide the control system with sensory input, which in turn triggers the motion of some other parts of the body (e.g., wheels) (cf. also Sharkey and Ziemke, 1998; Ziemke and Sharkey, 2001). Thus, much current robotics research, although to some degree acknowledging the role of the physical body, is still largely 'stuck' in the old distinction between hardware and software, which was central to computationalism and traditional AI. Hence, Searle's (1980) Chinese Room Argument (cf. Section 2) does in fact still apply to most of today's work in robotic AI (for the detailed argument see Ziemke, 2001a).

In summary, it can be said that, despite all biological inspiration, today's adaptive robots are still radically different from living organisms. In particular, despite their capacity for a certain degree of self-organization, today's so-called 'autonomous' agents are actually far from possessing the autonomy, and consequently the embodiment of living organisms. Mostly, this is due to the fact that today's robots typically are composed of mechanical parts (hardware) and computational control programs (software). The autonomy and subjectivity of living systems, on the other hand, emerges from the interaction of their components, i.e. autonomous cellular unities. Meaningful interaction between these first-order unities, and between the resulting second-order unity (the body) and its environment, is a result of their structural congruence, as pointed out by von Uexküll as well as Maturana and Varela. Thus, autonomy is a property of a living organism's organization right from its beginning as a cellular unity, and initial structural congruence with its environment results from the specific circumstances of reproduction. Its ontogeny maintains these properties throughout its lifetime through structural coupling with its environment.

Providing artifacts with the capacity for self-organization can be seen as the attempt to provide them with an *artificial ontogeny*. However, the attempt to provide them with autonomy this way seems to be doomed to fail, since it follows from the above argument that autonomy cannot from the outside be 'put' into a system, that does not already 'contain' it. Ontogeny preserves the autonomy of an organization, it does not 'construct' it. The attempt to bring the artifact into some form of structural congruence with its environment, on the other hand, can 'succeed', but only in the sense that the criterion for congruence cannot lie in the heteronomous artefact itself, but must be in the eye of the observer. This is exactly what happens when a robot is trained or evolved to adapt its structure in order to solve a task defined by its designer. Hence, using current technology, organismic embodiment is in fact limited to biological living systems.

5. Summary & Conclusion

This paper has discussed a number of diverse notions of embodiment. The motivation has been similar to that of Wilson (submitted), i.e. to disentangle the different claims and notions in the field. Unlike Wilson's paper, we have here focused on different notions of embodiment, i.e. the question exactly what kind of body is considered to be capable of embodied cognition. The notions we have identified in the literature are the following:

- *structural coupling* between agent and environment,
- *historical embodiment* as a result of a history of agent-environment interaction
- *physical embodiment*,
- '*organismoid' embodiment*, i.e. organism-like bodily form, and
- *organismic embodiment* of autopoietic, living systems

Roughly these notions can be considered as increasingly more restrictive, in the sense illustrated in Figure 1: organismic embodiment is a special case of organismoid embodiment and historical embodiment. Organismoid embodiment in turn is a special case of physical embodiment. Finally, both physical embodiment and historical embodiment are special cases of structural coupling.

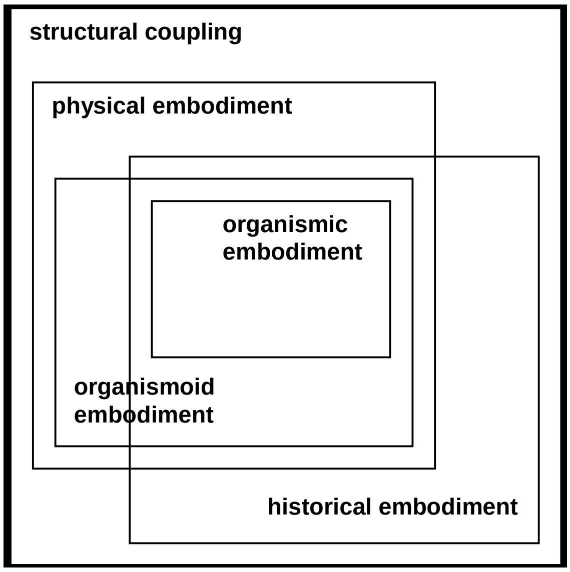

Figure 1: Notions of embodiment and their interrelations. Organismic embodiment is the most restrictive notion discussed in this paper, structural coupling the least restrictive.

In conclusion, despite much recent progress in adaptive/cognitive/epigenetic robotics the critique of Searle still holds (cf. Sharkey & Ziemke, in press; Ziemke, 2001a), i.e. the problem of the lack of intrinsic meaning or intentionality (in Searle's sense) in robots remains unsolved. That means, robotic AI might turn out to be a very powerful tool for the modeling of embodied cognition in the sense of a 'weak AI' (cf. Section 2). Using current technology, however, robots cannot themselves be embodied cognizers or minds in the sense of a 'strong AI'.

Obviously, this short paper has its limitations: ‘Social embodiment’, or the role of embodiment in social interactions, has not been addressed here at all; for a detailed discussion see Dautenhahn (1997). The most ‘restrictive’ or specific notion of embodiment discussed here, that of organismic embodiment, does in fact apply to all living systems, which is not particularly restrictive at all. Humanoid and human embodiment could be considered more restrictive, special cases of organismoid and organismic embodiment respectively. These might be considered to be of particular interest to cognitive science, but no arguments have been presented here as to why the more specific cases could or should allow for substantially different types of embodied cognition than other members of the more general categories. Nevertheless, we hope that the distinctions presented here will help to disentangle the large variety of claims, notions and theories that currently characterizes research on embodied cognition.

Acknowledgments

The author would like to thank Noel Sharkey, Henrik Svensson, and Stefan Berglund for a number of discussions that have contributed much to this paper. The author is supported by a grant (1507/97) from the Knowledge Foundation, Stockholm.

References

Brooks, R. A. (1990). Elephants don’t play chess. *Robotics and Autonomous Systems*, 6(1-2), 1-16.

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

Brooks, R. A. & Stein, L. A. (1993). *Building Brains for Bodies* (A.I. Memo No. 1439). Cambridge, MA: MIT, Artificial Intelligence Laboratory.

Brooks, R. A.; Breazeal, C.; Marjanovic, M.; Scasselati, B. & Williamson, M. (1998). The Cog Project: Building a Humanoid Robot. In C. Nehaniv (Ed.), *Computation for metaphors, analogy, and agents*, pages 52-87. New York: Springer.

Cariani, P. (1992). Some epistemological implications of devices which construct their own sensors and effectors. In Bourgine & Varela (Eds.), *Toward a practice of autonomous systems – Proceedings of the First European Conference on Artificial Life* (pp. 484-493). Cambridge, MA: MIT Press.

Clark, A. (1997). *Being There - Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Science, 9*, 345-351.

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

Dreyfus, H. L. (1996). The Current Relevance of Merleau-Ponty’s Phenomenology of Embodiment. *The Electronic Journal of Analytic Philosophy*, 4.

Duffy, B. R. (2001). *The Social Robot*. Doctoral dissertation, Dept. of Computer Science, University College Dublin, Ireland.

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

Franklin, S. A. (1997) Autonomous agents as embodied AI. *Cybernetics and Systems*, 25(8), 499-520.

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

Keijzer, F. (1998). Armchair consideration about wheeled behavior. In R. Pfeifer *et al.* (Eds.), *From Animals to Animats 5 – Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior* (pp. 13-21). Cambridge, MA: MIT Press.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*. 18, 513-549.

Lakoff, G. (1988). Cognitive Semantics. In U. Eco *et al.* (Eds.), *Meaning and Mental Representations*. Bloomington: Indiana University Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.

Lemmen, R. (1996). Embodiment for Non-Cartesians. *Conference of the European Society of Philosophy and Psychology*, Barcelona. Available as Cognitive Science Research Paper 441, COGS, University of Sussex, Brighton, UK.

Lipson, H. & Pollack, J. B. (2000). Automatic design and manufacture of robotic lifeforms. *Nature*, 406, 974-978 .

Lund, H. H.; Webb, B. & Hallam, J. (1998). Physical and temporal scaling considerations in a robot model of cricket calling song preference. *Artificial Life*, 4, 95-107.

Maturana, H. R. & Varela, F. J. (1980). *Autopoiesis and Cognition - The Realization of the Living*. Dordrecht, The Netherlands: D. Reidel Publishing.

Maturana, H. R. & Varela, F. J. (1987). *The Tree of Knowledge - The Biological Roots of Human Understanding*. Boston, MA: Shambhala.

Quick, T. & Dautenhahn, K. (1999). Making embodiment measurable. *Proceedings of ‘4. Fachtagung der Gesellschaft für Kognitions-wissenschaft’*. Bielefeld, Germany.

Quick, T.; Dautenhahn, K.; Nehaniv, C. & Roberts, G. (1999). On Bots and Bacteria: Ontology Independent Embodiment. In D. Floreano *et al.* (Eds.), *Proceedings of the Fifth European Conference on Artificial Life.* Heidelberg: Springer.

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

Prem, E. (1997). Epistemic autonomy in models of living systems. In *Proceedings of the Fourth European Conference on Artificial Life* (pp. 2-9). Cambridge, MA: MIT Press.

Riegler, A. (in press). When is a Cognitive System Embodied? *Cognitive Systems Research*, to appear.

Sharkey, N. E. & Ziemke, T. (1998). A consideration of the biological and psychological foundations of autonomous robotics. *Connection Science*, 10(3-4), 361-391.

Sharkey, N. E. & Ziemke, T. (2000). Life, Mind and Robots - The Ins and Outs of Embodied Cognition. In S. Wermter & R. Sun (Eds.), *Hybrid Neural Systems*. Heidelberg, Germany: Springer Verlag.

Sharkey, N. E. & Ziemke, T. (in press). Mechanistic vs. Phenomenal Embodiment - Can Robot Embodiment Lead to Strong AI? *Cognitive Systems Research*, to appear.

Stewart, J. (1996). Cognition = Life: Implications for higher-level cognition. *Behavioral Processes, 35,* 311-326.

Varela, F. J.; Thompson, E. & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

von Uexküll, J. (1909). *Umwelt und Innenwelt der Tiere*. Berlin: Springer Verlag.

von Uexküll, J. (1928). *Theoretische Biologie*. Berlin: Springer Verlag.

von Uexküll, J. (1982). The Theory of Meaning. *Semiotica*, 42(1), 25-82.

von Uexküll, J. (1985). Environment [Umwelt] and Inner World of Animals. In G. M. Burghardt (Ed.), *Foundations of Comparative Ethology*. New York: Van Nostrand Reinhold.

von Uexküll, T.; Geigges, W. & Herrmann, J. M. (1993). Endosemiosis. *Semiotica*, 96(1-2), 5-51.

Wilson, M. (submitted). Six views of embodied cognition. Manuscript submitted for journal publication. University of California, Santa Cruz.

Ziemke, T. (1999). Rethinking Grounding. In A. Riegler, M. Peschl & A. von Stein (Eds.), *Understanding Representation in the Cognitive Sciences*, pages 177-190. Plenum Press: New York.

Ziemke, T. (2000). *Situated Neuro-Robotics and Interactive Cognition*. Doctoral dissertation, Dept. of Computer Science, University of Sheffield, UK.

Ziemke, T. (2001a). The Construction of 'Reality' in the Robot: Constructivist Perspectives on Situated Artificial Intelligence and Adaptive Robotics. *Foundations of Science*, 6(1), 163-233.

Ziemke, T. (2001b). Disentangling Notions of Embodiment. *Workshop on Developmental Embodied Cognition*, Edinburgh, July 31, 2001.

Ziemke, T. & Sharkey, N. E. (2001). A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica*, 134(1-4), 701-746.

Zlatev, J. (1997). *Situated Embodiment. Studies in the Emergence of Spatial Meaning*. Doctoral dissertation, Dept. of Linguistics, Stockholm University, Sweden.

Zlatev, J. (2001). The epigenesis of meaning in human beings, and possibly in robots. *Minds and Machines*, *11*, 155-195.