Life, Mind, and Robots
The Ins and Outs of Embodied Cognition

Noel Sharkey[1] and Tom Ziemke[2,1]

[1] University of Sheffield
Dept. of Computer Science
Sheffield S1 4DP, UK
noel@dcs.shef.ac.uk
[2] University of Skövde
Dept. of Computer Science
54128 Skövde, Sweden
tom@ida.his.se

Abstract. Many believe that *the* major problem facing traditional artificial intelligence (and the functional theory of mind) is how to connect intelligence to the outside world. Some turned to robotic functionalism and a *hybrid response*, that attempts to rescue symbolic functionalism by grounding the symbol system with a connectionist hook to the world. Others turned to an alternative approach, *embodied cognition*, that emerged from an older tradition in biology, ethology, and behavioural modelling. Both approaches are contrasted here before a detailed exploration of embodiment is conducted. In particular we ask whether strong embodiment is possible for robotics, i.e. are robot "minds" similar to animal minds, or is the role of robotics to provide a tool for scientific exploration, a weak embodiment? We define two types of embodiment, Loebian and Uexküllian, that express two different views of the relation between body, mind and behaviour. It is argued that *strong* embodiment, either Loebian or Uexküllian, is not possible for present day robotics. However, weak embodiment is still a useful way forward.

1 Introduction

Cognitive science and artificial intelligence (AI) have been dominated by computationalism and functionalist theories of mind since their inception. The development of computers and their quick increase in information processing power during the 1940s and 50s led many theorists to assert that the relation between brain/body and mind in humans was the same as or similar to the relation between hardware and software in computers. The functionalist theory of mind seemed to solve in an elegant fashion the dispute between dualists and materialists on the relation between mind and matter. Thus, cognitive science and AI emerged as new disciplines, both of them initially based on the *computer metaphor for mind*.

S. Wermter and R. Sun (Eds.): Hybrid Neural Systems, LNAI 1778, pp. 313–332, 2000.

Serious doubts about the appropriateness of the computer metaphor were formulated in the arguments of Dreyfus [17] and Searle [37] around 1980, although for a long time they left most cognitive scientists and AI researchers cold. More attention was paid to the re-emergence of connectionism during the 1980s which, although to some degree concerned with neural hardware and thus taking a subsymbolic view on representation, did not question the computationalist framework in general. Ignoring the problems of purely computational theories of mind did not, however, make them disappear and theorists began to respond to the AI "mind in a vacuum" problem in one of two main approaches.

Firstly, there was the approach of those, such as Harnad [22], who recast the problem as the *symbol grounding problem* and suggested it could be solved by using a hybrid combination of the respective strengths of symbolic and connectionist theories. This was an attempt to maintain a computational theory of cognition augmented with *robotic* capacities supposed to connect internal representations to the world they are supposed to represent. The relationship between functional theories of mind, AI, connectionism, and the hybrid response is examined in Section 2. We focus on the separation of functionalism from the world and why this is seen as a weakness by opponents of strong AI. This leads us to a discussion of theoretical attempts to connect the physical symbol system to the world using connectionist nets as a front end.

The second approach, which is the main focus of this chapter, represents a more radical conception of cognition and AI arising out of biology and the study of animal behaviour. Led by a 'new robotics' that rejected traditional AI and representationalism altogether, it embraced the tenets of a *situated* and *embodied* cognition and challenged core assumptions of the funtionalist theory of mind. Two different types of embodiment are defined and discussed in Section 3 in relation to their biological underpinnings: *Loebian* or mechanistic embodiment and *Uexküllian* or phenomenal embodiment. Section 4 then moves on to critically examine two questions. Firstly, to what extent can the concept of embodiment be applied to artefacts such as robots? The arguments revolve around distinctions between living systems and machines that are made by humans, e.g. autopoiesis versus allopoiesis. Secondly, to what extent can *meaning* be attributed to the behaviour of a robot? It is argued that *strong* embodiment, either Loebian or Uexküllian, is not possible for present day robotics. However, weak embodiment is still a useful way forward.

2 Connectionism, Strong AI, and the Hybrid Response

In this section, we discuss some of the criticisms faced by proponents of the functional theory of mind and computationalism as instantiated in strong AI. Searle [38] states that the thesis of strong AI is that, "*the implemented program, by itself, is constitutive of having a mind. The implemented program, by itself, guarantees mental life*". The functional theory of mind is the view that mental states are physical states that are defined as functional states because of their causal relations. Thus, the functional role that a representation plays in a given computational system as a whole, imbues it with meaning (hence Function Role Semantics (FRS), [3]).

Connectionism is an attempt to get closer to the physical basis of mind by viewing representations as brain states. However, as Smolensky (1988) [50] pointed out, the subsymbolic representations of connectionists are closer to the symbolic realm than to physical brain states. The main debates between the cognitivists and the connectionists were about the form that representation should take, and how they are structured in order to realise their functional roles [44]. The cognitivists stood for syntactically structured compositional representations [18]. For them it was the syntactic relations between the elements of the representational system as a whole that allowed for systematic inference. In contrast, the connectionists argued for non-concatenative subsymbolic representations that were spatially rather than syntactically structured [45]. Several connectionists argued this case by building connectionist models that demonstrated systematicity of subsymbolic representation in some form or other [12,14,31,32].

So far as we have stated it, the debate was one within the remit of the functional theory of mind. But the connectionists also had mathematical learning techniques that were based on abstract models of neural computation. Thus the representations and their spatial organisation could be learned "from the world" by extensional programming, self-organisation, reinforcement learning, or, more lately, genetic algorithms. Although most of the representation research was carried out in worlds that mainly consisted of symbols of one form or another (but see [40]), the theory was that connectionist networks were hardwired physical machines (like brains) with connections that change in response to the outside world [39]. Connectionism had opened questions about the physical basis of representation and how neural adaptation can create and change representations and their spatial organisation through interaction with the world.

There were two ways in which the more philosophical connectionists could jump. One way was to go with eliminative materialists and see connectionism as the reduction of mental states to physical states, e.g. patterns of brain activation. It is here that the connectionists separate from 'run of the mill' cognitivists. It does not matter what machine the cognitivist runs the mind program on as long as it is capable of running it. For the connectionists, particular hardware was required, the brain. The other way to jump was to go for what Searle refers to as "causal emergence" [38], i.e. cognition emerges from a brain, [47].

There may be another type of strong AI lurking in these two positions. Stated in its most general form, if we were to physically implement a neural net system that had the causal powers of a brain, it would have mental states in the same way that we do. This does not sound much like strong AI. However, it could be argued that the properties of an artificial neural net are sufficient to provide a system with the causal power of a brain. Or, alternatively, the states of a neural net are sufficient to represent states of mind.

There were also those who did not want to 'go the whole hog' with connectionism and, instead, decided to occupy the middle ground by proposing hybrid systems that attempted to integrate the best features of both connectionism and symbolic computation (see [44]). Some of these researchers simply wanted to bring connectionism into the mainstream by embedding artificial neural networks in larger systems. Others, with commitments to some of the tenets of cognitivism, used the connectionist learning to "hook" symbols onto the world [22] (see also [46,61]), in an attempt to rescue functionalism from what they saw as one of its major weaknesses, i.e. minds are semantic, AI programs are purely syntactic, therefore AI programs cannot be minds [37].

Lloyd (1989) points out that a problem with such an internalist conception is that: "...some kernel of the system must win representationality by other means." [26]. He questions how a person might learn the meaning of a new word, say aham, without ostension, simply by using the context in which it occurs, just as is postulated to occur in a functional role semantics. This might be possible for a few terms, Lloyd argues, but there must come a critical point where there are just too many new terms to be able to allow an individual to learn them all without ostension (and to be able to provide ostensive definitions).

Lloyd's argument has similarities with John Searle's famous 1980 *Chinese Room argument* [37]. In brief, Searle imagines himself placed in a room where questions written in Chinese are passed to him and he uses a sort of lookup table combined with transformation rules to write down the correct answer in Chinese and pass it out of the room. The point is that Searle does not understand Chinese; he has no idea of the meaning relation between the questions and the answers; it is just a matter of manipulating meaningless symbols. Searle's point was that the operation of AI programs resembles what goes on in the Chinese room. Although the running programs manipulate symbols and provide outputs that are appropriate for their inputs, the programs or the machine do not understand what they are processing. There is no intrinsic meaning; the only meaning is for external observers. Therefore, AI programs are not intelligent in the strong sense.

Harnad (1990), taking Searle's criticisms on board, suggested a way to save strong AI. In what is termed a *Hybrid response*, the Chinese Room argument was recast as an argument against *symbolic functionalism* in which mind is held to be intrinsically, and solely, symbolic. Harnad proposed a *robotic functionalism* that extends symbolic functionalism to include some components that are intrinsically robotic. Robotic capacity, in this context, is a label for a range of sensorimotor interactions with the external environment, of which Harnad cites the discrimination and identification of objects, events, and states of affairs in the world as principal [22].

Harnad's view appears to be that sensory transduction is sufficient to foil the thrust of the Chinese Room argument. Harnad is quite explicit about the nature of the sensory transduction, or robotic capacity, that serves to ground a symbolic capacity: Specifically he names the ability to discriminate and identify nonsymbolic representations. Discrimination involves the machine making similarity judgements about pairs of inputs, and constructing *iconic* representations. These icons are nonsymbolic transforms of distal stimuli. However, Harnad argues that merely being able to discriminate inputs does not make for a robotic capacity.

In addition, the machine must also be able to identify the iconic representations. That is, it must be able to extract those invariant features of the sensory icons that will serve as a basis for reliably distinguishing one icon from another and for forming categorical representations.

But Searle is not just arguing that AI programs need a way to point at and categorise objects in the world. While Harnad sees the need to link the physical symbol system to the world to give the symbols a referential meaning in the same way as Lloyd *op cit.* has argued, he is still committed to functionalism. Searle, on the other hand, is arguing for much more; a machine cannot have intrinsic semantics because it is not intentional and it has no consciousness. To get the idea of what he means by consciousness, as opposed to the functionalist's representational states, Searle [38] asks us to pinch our skin sharply on the left forearm. He then lists a number of different things that will happen as a result and includes a list of neural and brain events. However, the most important event for Searle happens a few hundred milliseconds after you pinched your skin: "A second sort of thing happened, one that you know about without professional assistance. You felt pain. ... This unpleasant sensation had a certain particular sort of subjective feel to it, a feel which is accessible to you in a way that is not accessible to others around you." What he is getting at here are is that the feeling of pain is one of the many qualia. But Searle is not a dualist, he asks, "How is it possible for physical, quantitatively describable neuron firings to cause qualitative, private subjective experiences?". Searle's view is biological. He holds that the phenomenal mind is *caused by* a real living brain.

3 Embodied Cognition

Behaviour-based robotics represents an alternative approach to AI that has been gaining ground over the last decade (for overviews see [11,1,60,41]) . Although the basis for the approach has been around in biology for nearly a century and in robotics for over fifty years, it entered mainstream AI in the 1980s through the work and ideas of Rodney Brooks [6,7,8,9,10,11]. His approach to the study of intelligence was through the construction of physical robots embedded in, and interacting with, their environment by means of a number of *behavioural modules* working in parallel. The modules receive input from the sensors on the robot and the behaviour is controlled by the appropriate module for the situation. But there is no central controller. The modules are connected to each other in a way that allows certain modules to subsume the activity of others, hence this type of architecture is referred to as *subsumption architecture*.

This approach differs from both symbolic and robotic functionalism. Unlike classical AI, the modules do not make use of explicit internal representations in the sense of representational entities that correspond to entities in the 'real' world. Here, there is no need to talk about grounding symbols or representations in the world[1]. There are no internal representations to ground and there are

[1] Brooks is not entirely consistent with this point as he does write about about physical grounding of representations in [7,10].

no world models or planners; the world is its own world model for the robot. Intelligence is found in the interaction of the robot with its environment.

In terms of theory of mind, the literature appears to suggest a re-emergence of behaviourism in that there are no mental representations, just behaviour and predispositions to behaviour; sensing and acting in the world are necessary and sufficient for intelligence. Such a view takes form from considering a less anthropocentric universe than cognitivism. Much recent behaviour-based robotics takes inspiration from simple intelligent life forms such as arthropods, e.g. [2,57,25]. This provides a way to approach the fundamental building blocks of intelligence through a consideration of its biological basis.

But there is an unexpected turn in the theoretical picture that we have been painting about the new AI. Brooks (1991) proposed that the two cornerstones of the new approach to Artificial Intelligence are *situatedness* and *embodiment* [8]. Situatedness means that the physical world directly influences the behaviour of the robot - it is embedded in the world and deals with the world as it is, rather than as an abstract description. Embodiment commits the research to robots rather than to computer programs, as the object of study. This is consistent with the anti-mentalist stance of the new robotics.

However, the way in which embodiment is discussed has a least two radically different interpretations within the theory of mind, and behaviour-based roboticists appear to move between them with impunity. In the next two subsections, we examine these contrasting positions in relation to the work of the biologists, Charles Sherrington (1857-1952), Jacques Loeb (1859-1924) and Jakob von Uexküll (1864-1944), who, in different ways, developed theories of a biological basis for behaviour.

3.1 Loebian Embodiment

The first type of embodiment follows from the mechanistic or behaviourist line that Brooks appeared to be following. In this view one would expect embodiment to mean that cognition is embodied in the mechanism itself. That is, cognition is embodied in the control architecture of a sensing and acting machine. There is nothing else. There is no cognitive apparatus separate from the mechanism itself. There is no need for symbol grounding in that there are no symbols to ground. It is the behaviour or the dispositions to behaviour that are grounded in the interaction between agent and environment. This is similar to notions from physicalism in which the physical states of a machine are considered to be its mental states, i.e. there is no subjectivity. Movement is "forced" by the environment. This form of robot control relates mostly to the work of Sherrington on reflexes [49] and Loeb on tropisms [27].

Sherrington [49] focused on the nervous system and how it constitutes the behaviour of the multicellular animal as a "social unit in the natural economy". He proposed that it was nervous reaction that produces an animal individual from a mere collection of organs. The elementary unit of integration and behaviour was the simple reflex consisting of three separable structures: an effector organ, a conducting nervous pathway leading to that organ, and a receptor to initiate the reaction. This is the *reflex arc* and it is this simple reflex which exhibits the first grade of coordination. However, Sherrington admitted that the simple reflex is most likely to be a purely abstract conception. Since, in his view, all parts of the nervous system are connected together, no part may react without affecting and being affected by other parts. Nonetheless, the idea of chains of reflexes to form coherent behaviour became part of mainstream psychological explanation beginning with the work of the Russian psychologist Dimitri Pavlov who extended the study of reflex integration into the realm of animal learning, classical conditioning [33].

Loeb [27] had no patience with physiology and argued against the possibility of expressing the conduct of the whole animal as the "algegraic sum of the reflexes of its isolated segments". His concern was with how the whole animal reacted in response to its environment. Like the later behaviourists, Loeb was interested in how the environment "forced" or determined the movement of the organism. He derived his theory of *tropisms* (directed movement towards or away from stimuli) from the earlier scientific study of plant life, e.g. the directed movement through geotropism [24] and phototropism [16]. Thus for Loeb, animals were cartesian puppets whose behaviour was determined by the environmental puppeteer (see also [43,59]). Loeb would certainly have been very interested in today's biologically inspired robotics and was quick to see the implications of the artificial heliotropic machine built by J.J. Hammond. Loeb claimed that construction of the heliotropic machine, which was based on his theories, supported his mechanistic conception of the volitional and instinctive actions of animals [27].

Loeb used *tropism*[2] rather than *taxis* to stress what he saw as the fundamental identity of the curvature movements of plants and the locomotion of animals in terms of *forced movement* enabled by body symmetry. Although his specific theory about animal symmetry eventually fell under the weight of counter experimental evidence, major parts of his general theory of animal behaviour were taken up by later biologists and psychologists using the term *taxis* for directed animal movement. Fraenkel and Gunn [19], sympathising with Loeb's stance on the objective study of animal behaviour, heralded behaviour-based robotics by proposing that the behaviour of many organisms can be explained as a combination of taxes working together and in opposition.

However, it was the ideas of both reflex and taxis that first inspired what is now called artificial life research. Grey Walter (1950, 1953) built two "electronic tortoises" in one of the earliest examples of a physical implementation of intelligent behaviour. Each of these electromechanical robots was equipped with two 'sense reflexes'; a very small artificial nervous system built from a minimum of miniature valves, relays, condensers, batteries and small electric motors, and these reflexes were operated from two 'receptors': one photoelectric cell, giving the tortoises sensitivity to light, and an electrical contact which served as a touch receptor. The artificial tortoises were attracted towards light of moderate intensity, repelled by obstacles, bright light and steep gradients, and never stood

[2] Nowadays the term reflex is reserved for movements that are not directed towards the source of stimulation whereas taxis and tropism are used to denote movements with respect to the source of stimulation.

still except when re-charging their batteries. They were attracted to the bright light of their hutch only when their batteries needed re-charging. Grey Walter made many claims from his observations of these robots which included saying that the 'tortoises', exhibited hunger, sought out goals, exhibited self-recognition and mutual recognition. He also carried out the first artificial research on classical conditioning with his CORA system [21]. Grey Walter's work combined and tested ideas from a mixture of Loeb's tropisms and Sherrington's reflexes. Although Loeb is not explicitly mentioned in the book, the influence is clear, not least from the terms positive and negative tropisms instead of taxes.

These same ideas turn up again in recent robotics research and form the basis of much of modern robotics and Alife work (see also [48]). There is also considerable activity aimed at specific biological modelling that continues the line of mechanistic modelling of animal behaviour started by Hammond's heliotrope machine. Sherrington believed that there was much more to mind than the execution of reflexes as we shall see later. Loeb is more representative of the anti-mentalistic mechanistic view of intelligence and thus we dub this view *Loebian embodiment.*

3.2 Uexküllian Embodiment

Brooks [8] also espouses quite a different type of embodiment saying that, "The robots have bodies and experience the world directly - their actions are part of a dynamic with the world and have immediate feedback on their own sensations". He was anxious to take on board von Uexküll's concept of *Merkwelt* or perceptual world according to which each animal species, with its own distinctly non-human sensor suites and body, has it own phenomenal world of perception [5,9]. This notion of embodied cognition has its roots in von Uexküll's idea of bringing together an organism's perceptual and motor worlds in its *Umwelt* (subjective or phenomenal world) and hence we call it *Uexküllian embodiment.*

Von Uexküll tried to capture the seemingly tailor-made fit or solidarity between the organism's body and its environment in his formulation of a *theoretical biology* [53], and his *Umwelt* theory [52,54,56]. Von Uexküll criticized the mechanistic doctrine "that all living beings are mere machines" for the reason that it overlooked the organism's subjective nature, which integrates the organism's components into a purposeful whole. Thus his view is to a large degree compatible with Sherrington and Loeb's ideas of the organism as an integrated unit of components interacting in solidarity among themselves and with the environment. However, he differed from them in suggesting a rudimentary non-anthropomorphic psychology in which subjectivity acts as an integrative mechanism for coherent action:

> We no longer regard animals as mere machines, but as subjects whose essential activity consists of perceiving and acting ... for all that a subject perceives becomes his perceptual world and all that he does, his effector world. Perceptual and effector worlds together form a closed unit, the *Umwelt.* [54]

Although he strongly contradicted purely mechanistic/materialistic conceptions of life, and in particular the work of Loeb (e.g. [55]), von Uexküll was not a dualist. He did not deny the material nature of the organism's components. For example, he discussed how a tick's reflexes are "elicited by objectively demonstrable physical or chemical stimuli" [54]. However, for von Uexküll, the organism's components are forged into a coherent unit that acts as a behavioural entity. It is a *subject* that, through functional embedding, forms a "systematic whole" with its Umwelt.

Similar ideas have begun to emerge in robotics as part of a new *enactive cognitive science* approach, [51], and there is broad support in the field of *embodied cognition* where there is a reassessment of the relevance of life and biological embodiment for the study of cognition and intelligent behaviour, e.g. [15,13,36,58,34,61,62]. Two of the principals of the new approach, Maturana and Varela [29,30], have proposed that cognition is first and foremost a biological phenomenon. For them, "all living systems are cognitive systems, and living as a process is a process of cognition" [29].

In this framework, cognition is viewed as *embodied action* by which Varela *et al.* [51] mean "...first, that cognition depends upon the kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological, and cultural context". Thus, this view, like that of von Uexküll, emphasises the organism's embedding in not only its physical environment, but also the context of its own phenomenal world (Umwelt), and the tight coupling between the two. In the words of Varela *et al.*, "cognition in its most encompassing sense consists in the enactment or bringing forth of a world by a viable history of structural coupling" [51].

This is very different from Loebian embodiment where the mechanisms underlying behaviour are themselves controlled by the environment and where the organism is a mere Cartesian puppet (cf. also [43,59]). What we have called *Uexküllian* embodiment is the notion of a phenomenal cognition as an intrinsic part of a living body; as a process of that body.

4 Weak but Not Strong Embodiment

In this section we take a closer look at the two types of embodiment proposed in Section 3. First, Uexküllian embodiment is revisited to ask in what sense can a robot be a subjective embodiment. Then Loebian embodiment is discussed in terms of mechanistic minds and observer errors. For both types of embodiment we inquire as to whether they exist in the strong sense of equating "robot minds" with animal minds or they exist only in the weak sense of scientific or engineering models for investigating or exploiting knowledge about living systems.

4.1 Uexküllian Embodiment Revisited

A critical question facing Loeb, Sherrington, and von Uexküll concerned how a large number of living cells could act together in an integrated manner to produce a unitary behaving organism; in other words, to create a unitary living body. Loeb's approach was entirely behaviouristic and anti-mentalistic while von Uexküll proposed a subjective world as the method of integration. Sherrington was slighly different. He worked on the nervous mechanisms of the reflexes as an integrative mechanism. But he also recognised the limitations of the non-subjective. Writing about a decerberate dog that was used in his research, Sherrington states that,

> ... it contains no social reactions. It evidences hunger by restlessness and brisker knee jerks; but it fails to recognize food as food: it shows no memory, it cannot be trained to learn... The mindless body reacts with the fatality of a multiple penny-in-the-slot machine, physical, and not psychical. [49]

There can be no Uexküllian embodiment on existing robots. Uexküllian em*bodi*ment requires a living *body*. It is the embodiment of cognition or Umwelt as living processes. A robot does not have a body in this sense. It is a collection of inanimate mechanisms and non-moving parts that form a loosely integrated physical entity; a robot body is more like a car body and not at all like a living body. According to von Uexküll there are principal differences between the construction of a mechanism and a living organism, see e.g. [55,62]. He uses the example of a pocket watch to illustrate how machines are centripetally constructed: The individual parts of the watch, such as its hands, springs, wheels, and cogs must be produced first, so that they may be added to a common centerpiece. In contrast, the construction of an animal starts centrifugally; animal organs grow outwards from single cells. Von Uexküll was clear about the machine vision of animal life:

> Now we might assume that an animal is nothing but a collection of perceptual and effector tools [like microphones and motor cars], connected by an integrating apparatus which though still a mechanism, is yet fit to carry on the life functions. This is indeed the position of all mechanistic theories, whether their analogies are in terms of rigid mechanics or more plastic dynamics. They brand animals as mere objects. The proponents of such theories forget that, from the first, they have overlooked the most important thing, the *subject* which uses the tools, perceives and functions with their aid. [54]

More recently, Maturana and Varela [29] have also made a distinction between the organisation of living systems and machines made by humans. They were not interested in listing the static properties of living systems or the properties of the components of such systems. Rather they were concerned with the problem of how a living system, such as a single cell, can create and maintain its own identity despite a continual flux of perturbations and the continual changing of its components through destruction and transformation. As Boden [4], points out, this is universally accepted as one of the core problems of biology and for Maturana and Varela it is *the* problem.

Maturana and Varela *ibid* attacked the problem by concentrating on the organisation of matter in living systems. In his essay, *Biology of Cognition* in [29], Maturana proposed that the organisation of a cell is circular because the components that specify it are also the very components whose production and maintenance the organisation secures. It is this circularity that must be maintained for the system to retain its identity as a living system. Maturana and Varela [29] described the maintenance of the circularity with the new term *autopoiesis*, meaning self- (*auto*) -creating, -making, or -producing (*poiesis*).

An autopoietic machine, such as a living system, is a special type of homeostatic machine for which the fundamental variable to be maintained constant is its own organisation. This is unlike regular homeostatic machines which typically maintain single variables, such as temperature or pressure. The *structure* of an autopoietic system is the concrete realisation of the actual components (all of their properties) and the actual relations between them. Its *organisation* 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. Although these definitions apply to many systems, such as social systems, that may also be autopoietic, living systems are physical in a particular way. Moreover, autopoiesis is held to be necessary and sufficient to characterise the organisation of living systems.

Living systems are not the same as machines made by humans as some of the mechanistic theories would suggest. In Maturana and Varela's formulation, machines made by humans, including cars and robots, are *allopoietic*. Unlike an autopoietic machine, the organisation 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 organisation of the machine. Thus the changes that an allopoietic machine goes through without losing its defining organisation are necessarily subordinated to the production of something different from itself. In other words, it is not truly autonomous. In contrast, a living system 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 truly autonomous.

To be a living autonomous entity requires a unity that distinguishes an organism from its environment. At the core of autopoiesis lies the *autonomy* of the individual cells, which Maturana and Varela refer to as "first-order autopoietic units", similar to what von Uexküll meant by the term *Zellautonome* for autonomous cellular unities [53]. The individual cells' solidarity which constitutes the organism as an integrated behavioural entity and "second-order autopoietic unit" is due to the fact that "the structural changes that each cell undergoes in its history of interactions with other cells are complementary to each other, within the constraints of their participation in the metacellular unity they comprise" [30].

The chemical, mechanical, and integrating mechanisms of living things are missing from robots. Consequently, there can be no notion of multicellular solidarity or even a notion of a cell in a current robot. Although some may argue that the messaging between sensors, controllers and actuators, is a primitive type of integration, this is very different from the dependency relationship between living cells in real neural networks. Artificial neural nets can be used as a 'stand-in' integrative mechanism between sensors and actuators. However, they are not themselves integrated into the body of the robot; most of the body is a container for the controller, a stand to hang the sensors on, and a box for the motors and wheels. There is no interconnectivity or cellular communication. In multicellular creatures, solidarity of different types of cells in the body is required for survival. Cells need oxygen and so living bodies need to breathe, they need nutrition and so bodies need to behave in a way that enables ingestion of appropriate nutrients.

Furthermore, like von Uexküll [53], Maturana and Varela point out that living systems, cannot be properly analyzed at the level of physics alone, but require a *biological phenomenology*:

> ... autopoietic unities specify biological phenomenology as the phenomenology proper to those unities with features distinct from physical phenomenology. This is so, not because autopoietic unities go against any aspect of physical phenomenology - since their molecular components must fulfill all physical laws - but because the phenomena they generate in functioning as autopoietic unities depend on their organization and the way this organization comes about, and not on the physical nature of their components (which only determine their space of existence). [30]

Boden [4] has also recently argued that current metal robots cannot be living systems, or, in her words, "strong artificial life". Although she goes along with much of Maturana and Varela's characterisation of the organisation of the living, she stops short of accepting that all living processes are cognitive processes. Instead, Boden found it sufficient to argue from the weaker, but more straighforward, position that biochemical metabolism is necessary for life. Robots, like Grey Walter's turtles, that can recharge their batteries do not have metabolic processing that involves closely interlocking biochemical processes. They only receive packets of energy that do not produce and maintain the robot's body. It may be possible to produce an artificial life form, writes Boden, but it will probably have to be a biochemical one.

Even unicellular organisms with no nervous system exhibit more bodily integration and intimacy with their environment than current robots. In no sense does a 'situated' and 'embodied' robot actually *experience* the world. Its 'experience' is no different than that of an electronic tape measure. An electronic tape measure uses sonar to take measurements of room dimensions and ceiling heights in houses. You can point it at a wall and get a readout of say 54.4 cm. This is similar to the sonar sensing used for robots. The output from the robot sensors is a voltage proportional to distance. We could mount two of these tape measures on the front of a three wheeled platform with motors for the two rear wheels. The motors could be controlled directly by the tape measure outputs by wiring the left sensor to the right motor and vice versa. After a little tinkering, this device should make a reasonable job of avoiding boxes and walls. It is thus a situated robot in the sense described above. However, it does not have its "own sensations" nor does it have "a body" in the sense discussed above that could enable it to "experience the world directly". Thus, to repeat ourselves, strong Uexküllian embodiment is not possible on current robots.

Weak Uexküllian embodiment is, of course possible, in the sense of simulating embodied cognition with a physical robot. This would mean writing programs to capture aspects of cognition, but in a different way and with a different notion of cognition than used in disembodied AI. It would be a simulation of enactive cognition that could provide a 'wedge in the door' for biological and psychological research. In this sense it can be useful to view an autopoetic machine as an allopoietic machine. Although this has scientific value, according to Maturana and Varela, it will not reveal the autopoietic organisation.

4.2 Loebian Embodiment and the Clever Hans Error

Weak Loebian embodiment is not in question here. Robots have already been applied usefully as physical tests of the plausibility of mechanistic hypotheses about particular animal behaviour, e.g. [57,25]. Weak embodiment also incorporates the use of the robot as a thought tool for engineering or modelling applications [6,42].

However, why should strong embodiment be in question? After all, we have already said that the mechanistic or Loebian approach treats animals as mere Cartesian puppets. But they are puppets that have been co-evolved with their environments and their own niche that encapsulates the meaning of their existence. The intimate relationship between the body of a living organism and its environment as described in the previous section is missing, and this is important for strong embodiment regardless of one's views of mentality. Strong embodiment implies that the robot is integrated and connected to the world in the same way as an animal.

However, the identity of an allopoietic machine, like a robot, is not determined by its operation. Rather, because the products of the machine are independent of its organisation, its identity depends entirely on the observer. The meaning of the robot's "actions" is also in the observer's world and not in the "robot world". The robot's behaviour has meaning only for the observer. To think otherwise is to commit what is known in biology as the *Clever Hans error*. We shall expand on the nature and implications of such errors of attribution for the notion of strong Loebian embodiment in the remainder of this section.

Clever Hans was a horse whose performance startled citizens and scientists in the early part of the 20th century with its ability to perform simple arithmetic operations. A sum was written on a poster and displayed so that the horse could see it. Then the horse, to the amazement of the audience, would tap out the numerical solution with a hoof. There were even "scientific" theories about how it was using mental arithmetic. Hans even passed a test set up by Stumpf, the director of the Berlin Institute of Psychology, with a panel including a zoologist, a vet, and a circus trainer.

However, eventually the horse was submitted to an objective psychological assessment [35] and all was revealed. The horse always got the answer correct whenever there was an observer who knew the correct answer. With an observer who did not know the answer, Hans tapped an arbitrary number of times. It turned out that people who knew the answer to the sum were giving Hans a subtle cue at the right moment and he stopped hoofing. Such was Hans' ability at cue detection that it worked even if the observer knew what Hans was up to.

Hans did not know or care much about human arithmetic; it was not part of a horse's natural world. The exact number of hoof taps bore no relevance to Hans. If he responded appropriately to arbitrary start and stop cues, he was rewarded. Hans' cleverness was in being able to spot the stop cues without, apparently, having been trained. His owner, an Austrian aristocrat named von Osten, did not realise that he was being 'tricked'. Otherwise, the behaviour is not that much different to pigeons pecking in a Skinner box. It is as simple as that.

However, in the case of Clever Hans, it just so happens that, for human observers, the start sign, the sum, etc. have a ruleful (arithmetic) relationship with the behaviour of the horse. Thus the observers attributed causal significance to Hans' behaviour. Hans might have impressed them even more if they had given him questions such as, how many days are there in a week? Or, how many miles is it from Sheffield to Rotheram? Or, how many times has Brazil won the world cup? And, of course, he would also need a clever audience.

The point is that the meaning that the observer received from Clever Hans did not originate from the horse. Rather, it was merely a distorted reflection of the observer's own meaning. The meaning for Hans is less clear but it probably had nothing to do with arithmetic.

At first blush, the tale of Clever Hans seems to have similarities with Searle's Chinese room argument. Like Clever Hans, the person in the room was operating within the meaning domain of the observer. It was the observer who brought the meaning of the questioning and answering to the configuration. However, there is a major difference; Hans was simply responding to human cues to meet his own agenda. He was a living system as well. The observer was, in a sense, both in the 'room' of the system (as a cue provider) and outside of it at the same time (as the observer).

However, to invoke the systems argument, the most popular rebuttal of the Chinese Room argument, for Clever Hans would be inappropriate. The argument would run that although Hans did not know the meaning of the human arithmetic task, the system as a whole, consisting of the observer, the problem, Hans and his hoof, understood the human arithmetic problem. But this would be an oddly redundant system since one of its components, "the observer", can understand the whole problem so why have more components that understand nothing about it? The situation had an entirely different meaning for Hans that the observers did not understand. It would be a folly, however, to say that the 'system as a whole' understood the meaning of Hans' behaviour in terms of the horse world, for the same reasons as with the human observer.

Bringing the Clever Hans error to bear on robotics research, suppose that we have a light sensing robot, AutoHans, that has to perform the same task as its living counterpart. A screen with the sum on it is *lighted* and the robot begins moving back and forth (sadly there are no hoofs). When AutoHans has moved the correct number of times, the screen is *darkened* and AutoHans stops. These events are meaningful only to the observer and not to the machine. To think otherwise would be a Clever AutoHans error.

Going a step further, suppose that we now “rewire” the light sensing robot so that it can follow a light gradient. Now we can stand above it, call the sum out loud, and wave a torch around in circles until the correct number of circles has been achieved by the robot. This is such obvious trickery that it would not be worth mentioning except to highlight that it is no different than putting well placed lights in the environment to manipulate the behaviour of vehicles. For example, Grey Walter’s 1950s tortoise robots would be repelled by the bright light in their hutch (battery charger) and moved only towards moderate light in the room. When their batteries were low, they moved only towards the bright light in the hutch. This was a predesigned world and a robot with a carefully crafted controller. If this argument is sound, then to attribute “hunger”, “attraction”, or other anthropomorphic labels to the behaviour of these devices is again making the Clever AutoHans error. They can only exhibit weak Loebian embodiment.

Such systems, and this applies to all behavior-based robots, are designed by humans such that their movement in interaction with “cues” in the environment, e.g. lights of a particular intensity, looks, to human observers, like the behaviour of an organism. Even if the system adapts by neural and/or evolutionary methods, the goals or purpose of the system, by necessity, are designed by the researcher as the part that will make the work comprehensible to other human observers. These goals implicitly direct the themes of research papers and are what give the devices their credibility as autonomous agents. Essentially they search through “cue” space to find appropriate cues that lead to the satisfaction of the observer’s/experimenter’s goals. These goals are not the robot’s but the observer’s goals instantiated. Thus the robot’s interactions with the world carry meaning only for the observer.

Take for example, the case of Webb’s cricket studies [57,28] in which a wheeled robot uses an auditory system capable of selectively localising the sound of a male cricket stridulating (rubbing its wings together rapidly to produce a sound that attracts potential mates). Comparisons of the intensity of auditory information from two microphones was used to directly control the motors on a mobile robot and drive it towards the sound source.

Let us be clear about what this robot system tells us. It is a physical demonstration that mate selection can be mechanised through direct auditory-motor connections. No intermediate recognition or decision processes are necessary. This supports a biological model of mate selection as a taxis. Given the correspondence with some of the data on mate preferences in the male cricket, this is a useful piece of biological modelling. Sound localisation and its use in control is also useful from an engineering perspective.

However, care must be exercised not to attribute strong Loebian embodiment to the physical incantations of this robotic system. Webb is very careful with her claims. It is inappropriate, except perhaps in a popular science magazine headline, to call this a "robotic cricket". Yes, it can respond to certain sound frequencies and combinations of frequencies in a way that has some resemblances to the cricket; but only at a distal level of description and only for one taxis. This is at best a simulated partial embodiment.

Clearly at the proximal level of description there is little similarity between the real cricket and the robot behaviours. Obviously the instant-to-instant behavioural output of a rigid platform on wheels will be very different from a behavioural output of a legged insect body [23]. Both will behave quite differently towards a blade of grass on their trek to the male. The robot is actually a single function device; a tracker to find the most "attractive" male crickets inside a given sensory perimeter within a constrained physical environment. Out in the wild what chance would the robot have to locate the sound of a male cricket? And if it did, what are the chances that it would reach the male? (beware of mud patches, swampland, and rocks). Moving to the end game of mate selection, what if the robot did reach the male? What would the robot do when the cricket stopped stridulating? The robot would be a less than amorous companion. The cricket would certainly not recognise it as a mate even if he did notice it. Or, for that matter, what would the robot do if the cricket stridulated a few inches above it or even sat on top of it stridulating?

5 Conclusions

We began by examining one of the dominant criticisims that has dogged AI since the 1980s; namely that the symbolic realm of the AI program is connected to the world only through the knowledge of its human designers. Nor was connectionism free from this criticism although connectionist systems can be more readily connected to the outside world by "learning" a mapping between sensors and actuators. Indeed, many who agreed with the criticism but were still committed to the functional theory of mind developed a hybrid response to the problem by using connectionist nets to 'hook' the symbolic or representational domain onto the world

The same time, through the 1980s and 1990s, a new movement emerged in AI whose focus was primarily on robotics, i.e. physical machines which are, unlike computer programs, capable of interacting with their environment. Terms like "embodiment", "physical grounding" and "cognitive robotics" have become central in recent theories of mind, although they are used by cognitive theorists in at least two different ways. Firstly, as discussed in Section 2, there are those, who maintain the functionalist view of a computational mind and thus see the robot body as a means of "hooking" internal representations to the world. Secondly, as discussed in Section 3, there are those who reject the traditional notion of representation, and believe that cognition will emerge from the sensorimotor interaction of robot and environment.

We identified two significantly different notions of embodiment, Loebian (or mechanistic) and Uexküllian (or phenomenal), and discussed them in terms of their relationship to theories of body, mind and behaviour. In defining Loebian embodiment, we showed how the mechanistic notions of an early behaviourism, which pictures organisms as mechanisms more or less 'puppeteered' by their environment, have found their way into robotic studies of cognition and behaviour. This was contrasted with Uexküllian embodiment which encapsulates the view that living cognition relies on a phenomenal subjective interaction with the world. Cognition is 'embodied' in each organism's body and, in the extreme, all biological processes, including those of the single cell, are cognitive (e.g. Maturana and Varela [29,30]). However, because the new AI is still finding its (theoretical) feet, many of its practitioners move freely between these two notions of embodiment despite their differences.

It has been argued here that *strong* embodiment of either type is, in principle, not possible for current robots. Despite their apparent 'autonomy' during operation, robots remain allopoietic machines, which ultimately derive meaning only from their designers and observers. However, weak embodiment of both types is possible: There have been a number of successful robotic studies of mechanistic theories of animal behaviour. Similarly, robots can certainly be used to study and simulate how artificial agents can enact or bring forth, by means of adaptive techniques, their own 'phenomenal' environmental embedding in the form of interactive representations and behavioural structures. Thus, studying an allopoietic machine, such as a robot, *as if* it was physically autopoietic (living), can yield useful scientific insights and can bring much that is new into engineering. The limitations of such work, however, should be kept in mind. Two of those, dealt with here, are that, (i) current robots cannot experience a phenomenal world that arises directly out of having a living body, and (ii) the study of an allopoietic machine cannot reveal the organisation of the underlying autopoietic machine.

Acknowledgements

The authors would like to warmly thank Amanda Sharkey for helpful comments on an earlier draft of this paper.

References

1. Alan Arkin. *Behavior-Based Robotics.* Intelligent Robotics and Autonomous Agents Series. MIT Press, Cambridge, MA, 1998.
2. Randall D. Beer. *Intelligence as Adaptive Behavior - An Experiment in Computational Neuroethology.* Academic Press, San Diego, CA, 1990.
3. N. Block. Are absent qualia impossible? *Philosophical Review*, 89:257–272, 1980.
4. Margaret Boden. Is metabolism necessary? Technical Report CSPR 482, School of Cognitive and Computing Sciences, University of Sussex, Brighton, UK, January 1998.
5. Rodney A. Brooks. Achieving artificial intelligence through building robots. Technical Report Memo 899, MIT AI Lab, 1986.
6. Rodney A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2:14–23, 1986.

7. Rodney A. Brooks. Elephants don't play chess. *Robotics and Autonomous Systems*, 6(1–2):1–16, 1990.
8. Rodney A. Brooks. Intelligence Without Reason. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 569–595, San Mateo, CA, 1991. Morgan Kauffmann.
9. Rodney A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.
10. Rodney A. Brooks. The Engineering of Physical Grounding. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, pages 153–154, Hillsdale, NJ, 1993. Lawrence Erlbaum.
11. Rodney A. Brooks. *Cambrian Intelligence: The Early History of the New AI.* MIT Press, Cambridge. MA, 1999.
12. D.J. Chalmers. Syntactic transformations on distributed representations. *Connection Science*, 2:53–62, 1990.
13. Hillel J. Chiel and Randall A. Beer. The brain has a body: Adaptive behavior emerges from interactions of nervous system, body, and environment. *Trends in Neurosciences*, 20:553–557, 1997. Dec. 1997.
14. L Chrisman. Learning recursive distributed representations for holistic computation. *Connection Science*, 3:345–366, 1991.
15. Andy Clark. *Being There - Putting Brain, Body and World Together Again.* MIT Press, Cambridge, MA, 1997.
16. De Candolle. Reported in Frankel & Gunn (1940) *The Orientation of Animals: Kineses, Taxes and Compass Reactions*, Clarendon Press: Oxford, UK, 1832.
17. H.L. Dreyfus. *What computers can't do: The limits of artificial intelligence.* Harper & Row, New York, 2nd revised edition, 1979.
18. Jerry A. Fodor and Zenon W. Pylyshyn. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28:3–71, 1988.
19. G.S. Fraenkel and D.L. Gunn. *The Orientation of Animals: Kineses, Taxes and Compass Reactions.* Clarendon Press, Oxford, 1940.
20. William Grey Walter. An imitation of life. *Scientific American*, 182:42–54, 1950.
21. William Grey Walter. *The living brain.* Norton, New York, 1953.
22. Stevan Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.
23. Fred A. Keijzer. Some Armchair Worries about Wheeled Behavior. In *From animals to animats 5 - Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, pages 13–21, Cambridge, MA, 1998. MIT Press.
24. Knight. Reported in Frankel & Gunn (1940) *The Orientation of Animals: Kineses, Taxes and Compass Reactions*, Clarendon Press: Oxford, UK, 1806.
25. D. Lambrinos, M. Marinus, H. Kobayashi, T. Labhart, R. Pfeifer, and R. Wehner. An autonomous agent navigating with a polarized light compass. *Adaptive Behavior*, 6(1):131–161, 1997.
26. D. E. Lloyd. *Simple Minds.* MIT Press, Cambridge, MA, 1989.
27. Jacques Loeb. *Forced movements, tropisms, and animal conduct.* Lippincott Company, Philadelphia, 1918.
28. H.H. Lund, B. Webb, and J. Hallam. Physical and temporal scaling considerations in a robot model of cricket calling song preference. *Artificial Life*, 4(1):95–107, 1998.
29. H. R. Maturana and F. J. Varela. *Autopoiesis and Cognition - The Realization of the Living.* D. Reidel Publishing, Dordrecht, Holland, 1980.
30. H. R. Maturana and F. J. Varela. *The Tree of Knowledge - The Biological Roots of Human Understanding.* Shambhala, Boston, MA, 1987. Revised edition, 1992.

31. L. Niklasson and N.E. Sharkey. The systematicity and generalisation of connectionist compositional representations. In R.Trappl, editor, *Cybernetics and Systems.* Kluwer Academic Press, Dordrecht, NL, 1992.
32. L. Niklasson and T. van Gelder. On being systematically connectionist. *Mind and Language*, 3:288–302, 1994.
33. Ivan Petrovich Pavlov. *Conditioned Reflexes.* Oxford University Press, London, UK, 1927.
34. Rolf Pfeifer and Christian Scheier. *Understanding Intelligence.* MIT Press, Cambridge, MA, 1999.
35. P. Pfungst. *Clever Hans (The horse of Mr. von Osten): A contribution to experimental animal and human psychology.* Henry Holt, New York, 1911.
36. Erich Prem. Epistemic autonomy in models of living systems. In *Proceedings of the Fourth European Conference on Artificial Life*, pages 2–9, Cambridge, MA, 1997. MIT Press.
37. John Searle. Minds, brains and programs. *Behavioral and Brain Sciences*, 3:417–457, 1980.
38. J.R. Searle. *The Mystery of Consciousness.* Granta Books, London, 1997.
39. Noel E. Sharkey. Connectionist representation techniques. *Artificial Intelligence Review*, 5:143–167, 1991.
40. Noel E. Sharkey. Neural networks for coordination and control: The portability of experiential representations. *Robotics and Autonomous Systems*, 22(3-4):345–359, 1997.
41. Noel E. Sharkey. The new wave in robot learning. *Robotics and Autonomous Systems*, 22(3-4), 1997.
42. Noel E. Sharkey. Learning from innate behaviors: A quantitative evaluation of neural network controllers. *Autonomous Robots*, 5:317–334, 1998. Also appeared in *Machine Learning*, 31, 115-139.
43. Noel E. Sharkey and Jan Heemskerk. The neural mind and the robot. In A. J. Browne, editor, *Neural Network Perspectives on Cognition and Adaptive Robotics*, pages 169–194. IOP Press, Bristol, UK, 1997.
44. Noel E. Sharkey and Stuart A. Jackson. Three horns of the representational trilemma. In V. Honavar and L. Uhr, editors, *Symbol Processing and Connectionist Models for Artificial Intelligence and Cognitive Modeling: Steps towards Integration*, pages 155–189. Academic Press, Cambridge, MA, 1994.
45. Noel E. Sharkey and Stuart A. Jackson. An internal report for connectionists. In R. Sun and L. Bookman, editors, *Computational architectures integrating neural and symbolic processes: A perspective on the state of the art*, pages 223–244. Kluwer Academic Press, Boston, MA, 1995.
46. Noel E. Sharkey and Stuart A. Jackson. Grounding computational engines. *Artificial Intelligence Review*, 10(10):65–82, 1996.
47. Noel E. Sharkey and Amanda J.C. Sharkey. Emergent cognition. In J. Hendler, editor, *Handbook of Neuropsychology. Vol. 9: Computational Modeling of Cognition*, pages 347–360. Elsevier Science, Amsterdam, The Netherlands, 1994.
48. Noel E. Sharkey and Tom Ziemke. A consideration of the biological and psychological foundations of autonomous robotics. *Connection Science*, 10(3–4):361–391, 1998.
49. Charles Scott Sherrington. *The integrative action of the nervous system.* C. Scribner's Sons, New York, 1906.
50. P. Smolensky. On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11:1–74, 1988.

51. F. Varela, E. Thompson, and E. Rosch. *The Embodied Mind - Cognitive Science and Human Experience*. MIT Press, Cambridge, MA, 1991.
52. Jakob von Uexküll. *Umwelt und Innenwelt der Tiere*. Springer, Berlin, Germany, 1921.
53. Jakob von Uexküll. *Theoretische Biologie*. Suhrkamp, Frankfurt/Main, Germany, 1928.
54. Jakob von Uexküll. A stroll through the worlds of animals and men - a picture book of invisible worlds. In Claire H. Schiller, editor, *Instinctive Behavior - The Development of a Modern Concept*, pages 5–80. International Universities Press, New York, 1957. Originally appeared as von Uexküll (1934) *Streifzüge durch die Umwelten von Tieren und Menschen*. Springer, Berlin.
55. Jakob von Uexküll. The Theory of Meaning. *Semiotica*, 42(1):25–82, 1982.
56. Jakob von Uexküll. Environment [Umwelt] and inner world of animals. In G. M. Burghardt, editor, *Foundations of Comparative Ethology*, pages 222–245. Van Nostrand Reinhold, New York, 1985.
57. Barbara Webb. Using robots to model animals: A cricket test. *Robotics and Autonomous Systems*, 16(2–4):117–134, 1995.
58. Michael Wheeler. Cognition's coming home: The reunion of life and mind. In Phil Husbands and Inman Harvey, editors, *Proceedings of the Fourth European Conference on Artificial Life*, pages 10–19, Cambridge, MA, 1997. MIT Press.
59. Tom Ziemke. The 'Environmental Puppeteer' Revisited: A Connectionist Perspective on 'Autonomy'. In *Proceedings of the 6th European Workshop on Learning Robots (EWLR-6)*, pages 100–110, Brighton, UK, 1997.
60. Tom Ziemke. Adaptive Behavior in Autonomous Agents. *Presence*, 7(6):564–587, 1998.
61. Tom Ziemke. Rethinking Grounding. In Alexander Riegler, Markus Peschl, and Astrid von Stein, editors, *Understanding Representation in the Cognitive Sciences*. Plenum Press, New York, 1999.
62. Tom Ziemke and Noel E. Sharkey. A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica*, special issue on the work of Jakob von Uexküll, to appear in 2000.