A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life

TOM ZIEMKE and NOEL E. SHARKEY

Introduction

Much research in cognitive science, and in particular artificial intelligence (AI) and artificial life (ALife), has since the mid-1980s been devoted to the study of so-called *autonomous agents*. These are typically robotic systems situated in some environment and interacting with it using sensors and motors. Such systems are often self-organizing in the sense that they artificially learn, develop, and evolve in interaction with their environments, typically using computational learning techniques, such as artificial neural networks or evolutionary algorithms. Due to the biological inspiration and motivation underlying much of this research (cf. Sharkey and Ziemke 1998), autonomous agents are often referred to as 'artificial organisms', 'artificial life', 'animats' (short for 'artificial animals') (Wilson 1985), 'creatures' (Brooks 1990), or 'biorobots' (Ziemke and Sharkey 1998). These terms do not necessarily all mean exactly the same; some of them refer to physical robots only, whereas others include simple software simulations. But the terms all express the view that the mechanisms referred to are substantially different from conventional artifacts and that to some degree they are 'life-like' in that they share some of the properties of living organisms. Throughout this article this class of systems will be referred to as 'artificial organisms' or 'autonomous agents/robots' interchangeably.

The key issue addressed in this article concerns the semiotic status and relevance of such artificial organisms. The question is whether and to what extent they are autonomous and capable of semiosis. This is not straightforward since semiosis is often considered to necessarily involve living organisms. Morris (1946), for example, defines semiosis as 'a sign-process, that is, a process in which something is a sign to some organism'. Similarly, Jakob von Uexküll[1] considered signs to be 'of prime importance in all aspects of life processes' (T. von Uexküll 1992), and made a clear distinction between organisms, which as *autonomous subjects* respond to

Semiotica 134–1/4 (2001), 701–746 0037–1998/01/0134–0701

signs according to their own *specific energy*, and inorganic mechanisms, which lack that energy, and thus remain *heteronomous* (for a more detailed discussion see the following section).

Mechanisms can, of course, be involved in sign processes, in particular computers and computer programs.[2] They are, however, typically considered to lack 'first hand semantics', i.e., 'intrinsic meaning' (Harnad 1990) or 'contents for the machine' (Rylatt et al. 1998), and to derive their semantics from the fact that they are programmed, observed, and/or interpreted by humans. Andersen et al. (1997) have argued in detail that computers/programs, when it comes to semiosis, fall somewhere in between humans and conventional mechanisms, but that they ultimately derive their semiotic 'capacities' from the interpretation of their designers and users. The major difference, they argued, was that living systems are autopoietic, i.e., self-creating and -maintaining, whereas machines are not (this issue will be discussed in detail later). Hence, their 'tentative conclusion' was that

> … the difference between human and machine semiosis may not reside in the particular nature of any of them. Rather, it may consist in the condition that machine semiosis presupposes human semiosis and the genesis of the former can be explained by the latter. (Andersen et al. 1997: 569)

Cognitive science and AI research has, in fact, since its beginning in the 1950s been dominated by the so-called *computer metaphor for mind*, i.e., the view that the human mind is very much like a computer program. This has led decades of traditional AI research to fall into the *internalist trap* (Sharkey and Jackson 1994) of focusing solely on disembodied computer programs and internal *representations* supposed to 'mirror' a pre-given external reality (cf. Varela et al. 1991), while forgetting about the need for grounding and embedding these in the world they were actually supposed to represent. Hence, for cognitive scientists the use of embodied, situated agents offers an alternative, bottom-up approach to the study of intelligent behavior in general, and internal representation and sign usage in particular.

Artificial organisms, unlike computer programs equipped with robotic capacities of sensing and moving, *do* interact with their environments, and they appear to do so independently of interpretation through external users or observers. Moreover, such systems are often *self-organizing*, i.e., they 'learn', 'develop', and 'evolve' in interaction with their environments, often attempting to mimic biological processes. Several examples of this type of self-organization in artificial organisms will be discussed throughout this article. The sign processes and functional circles by which artificial organisms interact with their environments are, therefore,

typically self-organized, i.e., the result of adaptation in interaction with an environment, rather than programmed or built-in by a designer, and thus often not even interpretable to humans (cf. Prem 1995). Hence, unlike computer programs, their genesis typically cannot be explained with reference to human design and interpretation alone. Thus, it has been argued that autonomous agents are, at least in theory, capable of possessing 'first hand semantics' (e.g., Harnad 1990; Brooks 1991b; Franklin 1997; Bickhard 1998). Their semiotic and epistemological interest, it is held, arises because unlike conventional machines, their use of signs and representations is self-organized, and thus, as for living systems, largely private and typically only meaningful to themselves. Many researchers, therefore, no longer draw a strict line between animals and autonomous robots. Prem (1998), for example, refers to both categories as 'embodied autonomous systems', and does not at all distinguish between living and non-living in his discussion of semiosis in such systems. We have previously discussed this distinction in an examination of the biological and psychological foundations of modern autonomous robotics research (Sharkey and Ziemke 1998). In that article we investigated differences between the 'embodiment' of living and non-living systems, and their implications for the possibility of cognitive processes in artifacts. In this article the issues are further analyzed with reference to Jakob von Uexküll's theory of meaning.

As a result of the new orientation towards agent-environment interaction and biological inspiration, the work of Jakob von Uexküll by some researchers has been recognized as relevant to the study of robotics, ALife, and embodied cognition. Examples are the works of Brooks (1986a, 1991a); Emmeche (1990, 1992, this issue); Prem (1996, 1997, 1998); Clark (1997); and our own recent work (Sharkey and Ziemke 1998, 2000; Ziemke 1999b). However, a detailed analysis and discussion of Uexküll's theory, its relation to and implications for recent theories in AI and cognitive science, is still lacking; hence, this is what we aim to provide in this article. We believe that Uexküll's theory can contribute significantly to the field by deepening the understanding of the use of signs and representations in living beings and clarifying the possibilities and limitations of autonomy and semiosis in artificial organisms.

The scene is set in the next section in a discussion of the contrasting positions of Jacques Loeb and Jakob von Uexküll on the differences between organisms and mechanisms. This leads into a discussion of attempts by AI to endow mechanisms with some of the mental and behavioral capacities of living organisms. Moreover, the history of different approaches to AI is discussed with an emphasis on the connections to issues in semiotics, and in particular the relation to Uexküll's work.

The following section then takes this a step further by detailing the issues involved in the self-organization of artificial organisms through adaptation of sign processes using computational evolution and learning techniques. Then there will be a discussion of how artificial organisms interact with objects and other agents in their environment by means of sign processes, and how this distinguishes them from the conventional mechanisms discussed by Uexküll. In the penultimate section Uexküll's theory is compared to the closely related work of Maturana and Varela on the biology of cognition. Using both these theoretical frameworks, we further examine the role of the living body in the use of signs/representations. Finally, we consider the implications of not having a living body for the possibility and limitations of autonomy and semiosis in artificial organisms.

Organisms versus mechanisms

Many of the ideas discussed in modern autonomous robotics and ALife research can already be found in biological and psychological discussions from the late nineteenth and early twentieth century. Jacques Loeb (1859–1924) and Jakob von Uexküll (1864–1944) represented the discontent felt by a number of biologists about anthropomorphic explanations and they both were influential in developing a biological basis for the study of animal behavior, although in very different ways. After Darwin's (1859) book, *The Origin of Species*, comparative psychology had attempted to find a universal key which resulted in the breaking down of the distinction between humans and other species. This led to the attribution of human-like mental qualities to other vertebrates and even invertebrates. In stark contrast to this anthropomorphic approach, Loeb developed scientifically testable mechanistic theories about the interaction of organism and environment in the creation of behavior. Uexküll, on the other hand, theorized about organism-environment interaction in terms of subjective perceptual and effector worlds, and thus contradicted anthropomorphic as well as purely mechanistic explanations. What united Loeb and Uexküll was the goal to find a way to explain the behavioral unity of organisms, and their environmental embedding, based on their biology; in their individual approaches, however, they differed substantially.

Mechanistic theories

Loeb (1918) derived his theory of *tropisms* (directed movement towards or away from stimuli) by drawing lessons from the earlier scientific study

of plants where considerable progress had been made on directed movement through geotropism (movement with respect to gravity) (Knight 1806) and phototropism (movement with respect to light) (De Candolle 1832). Strasburger (1868) really set the ball rolling for animal behavior in a study of the movements of unicellular organisms towards light which he labelled *phototaxis* to distinguish the locomotory reactions of freely moving organisms from the *phototropic* reactions of sedentary plants. The study of *chemotaxis* came soon afterwards (e.g., Pfeffer 1883) to describe attractions of organisms to chemicals. Although Loeb wanted to explain the behavior of higher organisms, those with nervous systems, he continued to use the term *tropism* 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.

Umwelt and counterworld

Uexküll strongly criticized the purely mechanistic doctrine 'that all living beings are mere machines' (Uexküll 1957) in general, and Loeb's work in particular (e.g., Uexküll 1982), for the reason that it overlooked the organism's subjective nature, which integrates the organism's components into a purposeful whole. Thus, although his view is to some degree compatible with Loeb's idea of the organism as an integrated unit of components interacting in solidarity among themselves and with the environment, he differed from Loeb in suggesting a non-anthropomorphic psychology in which subjectivity acts as an integrative mechanism for agent-environment coherence.

> The mechanists have pieced together the sensory and motor organs of animals, like so many parts of a machine, ignoring their real functions of perceiving and acting, and have gone on to mechanize man himself. According to the behaviorists, man's own sensations and will are mere appearance, to be considered, if at all, only as disturbing static. But we who still hold that our sense organs serve our perceptions, and our motor organs our actions, see in animals as well not only the mechanical structure, but also the *operator, who is built into their organs as we are into our bodies*. We no longer regard animals as mere machines, but as subjects whose essential activity consists of perceiving and acting. We thus unlock the gates that lead to other realms, 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*. (Uexküll 1957: 6; first emphasis added)

Uexküll (1957) used the now famous example of the tick to illustrate his concept of *Umwelt* and his idea of the organism's embedding in its world

through *functional circles*.[3] It is three such functional circles in 'well-planned succession' which coordinate the interaction of the tick as a subject (and *meaning-utilizer*) and a mammal as its object (and *meaning-carrier*):

(1) The tick typically hangs motionless on bush branches. When a mammal passes by closely its skin glands carry perceptual meaning for the tick: the perceptual signs (*Merkzeichen*) of butyric acid are transformed into a perceptual cue (*Merkmal*) which triggers effector signs (*Wirkzeichen*) which are sent to the legs and make them let go so the tick drops onto the mammal, which in turn triggers the effector cue (*Wirkmal*) of shock.
(2) The tactile cue of hitting the mammal's hair makes the tick move around (to find its host's skin).
(3) The sensation of the skin's heat triggers the tick's boring response (to drink its host's blood).

Uexküll did not deny the physical/chemical nature of the organism's components and processes, i.e., his view should not, as is sometimes done, be considered vitalistic[4] (cf. Emmeche 1990, this issue; Langthaler 1992). He 'admitted' that the tick exhibits 'three successive reflexes' each of which is 'elicited by objectively demonstrable physical or chemical stimuli'. But he pointed out that the organism's components are forged together to form a coherent whole, i.e., a *subject*, that acts as a behavioral entity which, through functional embedding, forms a 'systematic whole' with its Umwelt.

> We are not concerned with the chemical stimulus of butyric acid, any more than with the mechanical stimulus (released by the hairs), or the temperature stimulus of the skin. We are solely concerned with the fact that, out of the hundreds of stimuli radiating from the qualities of the mammal's body, only three become the bearers of receptor cues for the tick. … What we are dealing with is not an exchange of forces between two objects, but the relations between a living subject and its object. … The whole rich world around the tick shrinks and changes into a scanty framework consisting, in essence, of three receptor cues and three effector cues — her *Umwelt*. But the very poverty of this world guarantees the unfailing certainty of her actions, and security is more important than wealth. (Uexküll 1957: 11f.)

As T. von Uexküll (1997b) pointed out, the model of the functional circle contains all the elements which are part of a sign process, and whose interaction forms the unity of a semiosis: an organism is the *subject* (or interpreter), certain environmental signals play the role of *signs* (or interpretanda), and the organism's biological condition determines

the *behavioral disposition* (or interpretant). The *object* (interpretatum), on the other hand, can be harder to identify using common sign-theoretic concepts, since for the organism, e.g., the tick, it does not necessarily exist as an abstract entity, e.g., 'a mammal', but might only have temporary existence as different *semiotic objects* and the bearer of varying meanings, e.g., three different ones in the tick's case. Hence, Uexküll sometimes referred to the sign processes in the nervous system as a 'mirrored world' (Uexküll 1985; cf. also T. von Uexküll et al. 1993), but pointed out that by that he meant a '*counterworld*', not a 1 : 1 reflection of the external environment. Thus, he wanted to emphasize that

> … in the nervous system the stimulus itself does not really appear but its place is taken by an entirely different process which has nothing at all to do with events in the outside world. This process can only serve as a *sign* which indicates that in the environment there is a stimulus which has hit the receptor but it does not give any evidence of the quality of the stimulus. (Uexküll 1909: 192)[5]

T. von Uexküll et al. (1993) also point out that the notion of 'counterworld' should not be equated with a 'mirror' in the narrow sense of a reflection of the environment. They further elaborate that

> … in this phenomenal universe [of the counterworld], the objects of the environment are represented by schemata which are not, as in a mirror, products of the environment, but rather 'tools of the brain' ready to come into operation if the appropriate stimuli are present in the outside world. In these schemata, sensory and motor processes are combined … to form complex programs controlling the meaning-utilizing … behavioral responses. They are retrieved when the sense organs have to attribute semiotic meanings to stimuli. (T. von Uexküll et al. 1993: 34)

Hence, T. von Uexküll (1992: 308) concludes that an 'essential problem, which he [Jakob von Uexküll] has solved through the model of a circular process, is the relationship between sign and behavior (perception and operation)'.

Autonomy

The key difference between mechanisms and living organisms is, according to Uexküll, the *autonomy* of the living. Following the work of Müller (1840), he pointed out that 'each living tissue differs from all machines in that it possesses a ''specific'' life-energy in addition to physical energy' (Uexküll 1982: 34), which allows it to react to different stimuli with a 'self-specific' activity according to its own 'ego-quality' (*Ich-Ton*), e.g., a muscle with contraction or the optic nerve with sensation of light. Hence, each

living cell *perceives* and *acts*, according to its specific perceptual or receptor signs and impulses or effector signs, and, thus, the organism's behaviors 'are not mechanically regulated, but meaningfully organized' (Uexküll 1982: 26). The operation of a machine, on the other hand, is purely mechanical and follows only the physical and chemical laws of cause and effect. Furthermore, Uexküll (1928: 180)[6] referred to Driesch, who pointed out that all action is a mapping between individual stimuli and effects, depending on a historically created basis of reaction (*Reaktionsbasis*), i.e., a context-dependent behavioral disposition (cf. Driesch 1931). Mechanisms, on the other hand, do not have such a historical basis of reaction, which, according to Uexküll, can only be grown — and there is no growth in machines. Uexküll (1928: 217) further elaborates that the rules machines follow are not capable of change, due to the fact that machines are fixed structures, and 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* (cf. also T. von Uexküll 1992). Machines can, therefore, 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* (Uexküll 1928: 301).

This is also closely related to what Uexküll described as the 'principal difference between the construction of a mechanism and a living organism', namely that 'the organs of living beings have an innate meaning-quality, in contrast to the parts of machine; therefore they can only develop centrifugally':

> 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. (Uexküll 1982: 40)

In a later section we will discuss in detail the relation between Uexküll's theory and Maturana and Varela's (1980, 1987) more recent work on autopoiesis and the biology of cognition.

Mechanistic and cybernetic models

Although Uexküll and others presented strong arguments against the mechanistic view, a number of researchers during the first half of the twentieth century began to build machines to test mechanistic hypotheses about the behavior of organisms. Beside the work of Loeb, inspiration was taken in particular from Sherrington's (1906) work on *reflexes* and even earlier work on *taxis* (see Fraenkel and Gunn 1940 for an overview of nineteenth-century research on different forms of taxis). Loeb (1918) himself described a heliotropic machine[7] constructed by J. J. Hammond and held that:

> … the actual construction of a heliotropic machine not only supports the mechanistic conception of the volitional and instinctive actions of animals but also the writer's theory of heliotropism, since the theory served as the basis in the construction of the machine. (Loeb 1918)

One of the most impressive early examples of research on artificial organisms came from Walter (1950, 1951, 1953), who built his two electronic tortoises, Elmer and Elsie, of the species *Machina speculatrix* between 1948 and 1950. Among other things, they exhibited phototaxis and 'hunger'; they re-entered their hutch to recharge their batteries as required. This work combines and tests ideas from a mixture of Loeb's tropisms and Sherrington's reflexes.[8] Although Loeb is not explicitly mentioned in the book, the influence is clear, not least from the terms positive and negative tropisms. Walter's electromechanical creatures were equipped with two 'sense reflexes'; a little 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. Elmer and Elsie were attracted towards light of moderate intensity, repelled by obstacles, bright light, and steep gradients, and never stood still except when re-charging their batteries. They were attracted to the bright light of their hutch only when their batteries needed re-charging. These archetypes of biologically inspired robotics exhibited a rich set of varying behaviors, including 'goal finding', 'self-recognition', and 'mutual recognition' (Walter 1953).

Although much of this work ran somewhat counter to Uexküll's sharp critique of the mechanistic doctrine, these early mechanistic and cybernetic attempts at building forms of what is now called ALife were, in their general technical conception, nevertheless, to some degree compatible

with his view of the interaction between organism and environment (cf. also Emmeche, this issue). In particular, organisms were modeled/constructed as embedded in their environment by means of functional circles, i.e., (seemingly) intelligent behavior was viewed as the outcome of a continual interaction between organism and environment in bringing forth effective behavior, and signs were viewed as playing a functional role in this interaction. This is not to say that there are no significant differences between Uexküll's and the mechanists' positions (as discussed above, of course there are), but as we will see in the following section, these two views are actually significantly closer to each other than either of them is to the approach to the study of intelligent behavior that most AI research was taking during the 1950–1980s, in particular its strict distinction and separation between internal representations and external world.

AI: From artificial organisms to computer programs and back

The endeavor of AI research can be characterized as the attempt to endow artifacts with some of the mental and behavioral capacities of living organisms. Thus, the early work on artificial organisms discussed in the previous section could be seen as a forerunner of the field of AI, which began to form under that name in the mid-1950s. Somewhat ironically, however, AI research almost completely ignored that early biologically motivated work for about thirty years. As we will see in the next subsection, AI researchers, initially focusing on mental capacities, turned to the computer as a model of mind instead. It was not until the mid-1980s that parts of the AI community returned to its roots and began to focus on behavior and agent-environment interaction again, as will be discussed in detail later.

A much debated concept in AI research and the other cognitive sciences has always been the notion of *representation* as the connection between agent and world. How exactly cognitive representation 'works', has been, as we will see in the following, a topic of controversy. Although the different notions of representation and their usage largely overlap with different semiotic notions of signs and their usage, semiotics has had relatively little direct impact on cognitive science and AI research. Unfortunately, there has been less interaction between the disciplines than one might expect given the common interest in signs and representations. We will here refer to signs and representations as roughly similar and interchangeable notions, and particularly focus on the development of the notion of representation in AI.

Cognitivism and the computer metaphor for mind

During the 1940s and 1950s a growing number of researchers, like Uexküll discontent with *behaviorism* as the predominant paradigm in the study of mind and behavior, became interested in the mind's internal processes and representations, whose study behaviorists had rejected as being unscientific. This revived the central idea of *cognitive psychology*, namely, that the brain possesses and processes information. This idea can be found in the much earlier work of William James (1892). Craik, however, in his 1943 book, *The Nature of Explanation*, was perhaps the first to suggest that organisms make use of explicit knowledge or world models, i.e., internal representations of the external world:

> If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it. (Craik 1943: 61)

Craik had little to say about the exact form of the internal representations or the processes manipulating them (cf. Johnson-Laird 1989). However, he was fairly specific about what he meant by a 'model', namely, something that is much closer to a 'mirror' of external reality than Uexküll's notion of a 'counterworld'.

> By a model we … mean any physical or chemical system which has a similar relation-structure to that of the processes it imitates. By 'relation-structure' I [mean] … the fact that it is a physical working model which works in the same way as the processes it parallels, in the aspects under consideration at any moment. (Craik 1943: 51)

At the same time computer technology became increasingly powerful. Researchers began to realize the information processing capabilities of computers and liken them to those of humans. Taken to extremes, this analogy echoes one of the central tenets of *cognitivism*, which considers cognition to be much like a computer program that could be run on any machine capable of running it. In this functionalist framework of the *computer metaphor for mind*, having a body, living or artificial, is regarded as a low-level implementational issue. Even connectionism of the 1980s, with its biologically inspired computation and its strong criticisms of the cognitivist stance for its lack of concern with neural hardware, was mainly concerned with explaining cognitive phenomena as separated from organism-world interaction.

Thus, the early work on the interaction between cybernetic/robotic organisms and their environments was divorced from the dominant themes in the mind sciences. The early biologically-oriented approaches contrasted sharply with those of cognitivism, traditional AI, and traditional cognitive psychology. Here, mind was cut off from body in a move that echoes in reverse the studies of decerebrated animals carried out by Sherrington (1906) and others. Neisser (1967), for example, in his book *Cognitive Psychology*, which defined the field, stressed that the cognitive psychologist 'wants to understand the program, not the hardware'. According to Neisser, 'the task of a psychologist trying to understand human cognition is analogous to that of a man trying to understand how a computer has been programmed'.

Hence, while behaviorists had treated mind as an opaque box in a transparent world, cognitivists treated it as a transparent box in an opaque world (Lloyd 1989). Research in cognitive science and AI, therefore, focused on what Uexküll referred to as the 'inner world of the subject' (Uexküll 1957). The cognitivist view, largely following Craik, is that this 'inner world' consists of an internal model of a pre-given 'external reality', i.e., *representations* (in particular symbols) corresponding/referring to external objects ('knowledge'), and the computational, i.e., formally defined and implementation-independent, processes operating on these representations ('thought'). That means, like Uexküll's theory, cognitivism was strictly opposed to behaviorism and emphasized the importance of the subject's 'inner world', but completely unlike Uexküll it de-emphasized, and in fact most of the time completely ignored, the environmental embedding through functional circles. Or in Craik's terms: cognitivism became pre-occupied with the internal 'small-scale model', and the idea that it was to be located 'in the head' alone, but completely neglected both organism and reality.

An example of the cognitivist *correspondence* notion of representation was given by Palmer (1978), who characterized a *representational system* as including the following five aspects: (1) the represented world, (2) the representing world, (3) what aspects of the represented world are being modeled, (4) what aspects of the representing world are doing the modeling, and (5) what the *correspondences* between the two worlds are. Thus, the cognitivist view of the relation between internal model and external world was as illustrated in Figure 1, i.e., representation was seen as internal mirror of an observer-independent, pre-given external reality (cf. also Varela et al. 1991).

During the 1970s traditional AI's notion of representation, as illustrated in Figure 1, came under attack. Dreyfus (1979) pointed out that AI programs represented descriptions of isolated domains of human

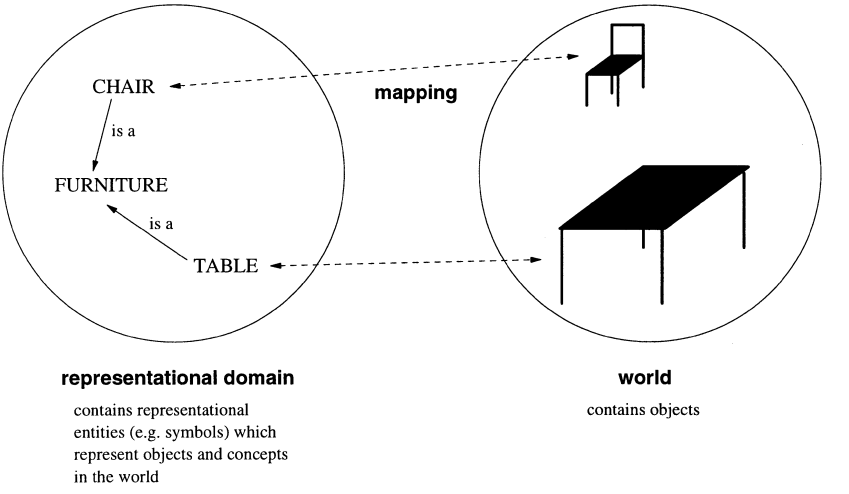

Figure 1. *The idea of traditional AI representation as a direct mapping between internal representational entities, e.g., symbols and objects in the external world. Adapted from Dorffner (1997)*

knowledge (‘*micro-worlds*’) ‘from the outside’. That means, they were not ‘situated’ in them due to the fact they always lacked a larger background of, e.g., bodily skills or cultural practices, which might not be formalizable at all. In a similar vein Searle (1980) pointed out that, because there are no causal connections between the internal symbols and the external world they are supposed to represent, purely computational AI systems lack *intentionality*.[9] In other words, AI systems do not have the capacity to relate their internal processes and representations to the external world. It can be said in semiotic terms that what AI researchers intended was that the AI system, just like humans or other organisms, would be the interpreter in a triadic structure of sign (internal representation/symbol), external object, and interpreter. What they missed out on, however, was that, due to the fact that, in Uexküll’s terms, the ‘inner world of the subject’ was completely cut off from the external world by traditional AI’s complete disregard for any environmental embedding through receptors and effectors, the interpreter could not possibly be the AI system itself. Hence, as illustrated in Figure 2, the connection or mapping between representational domain and represented world is really just in the eye (or better: the mind) of the designer or other observers.

The problem illustrated in Figure 2 is now commonly referred to as the *symbol grounding problem* (Harnad 1990). A number of other authors, however, have pointed out that the grounding problem is not limited to

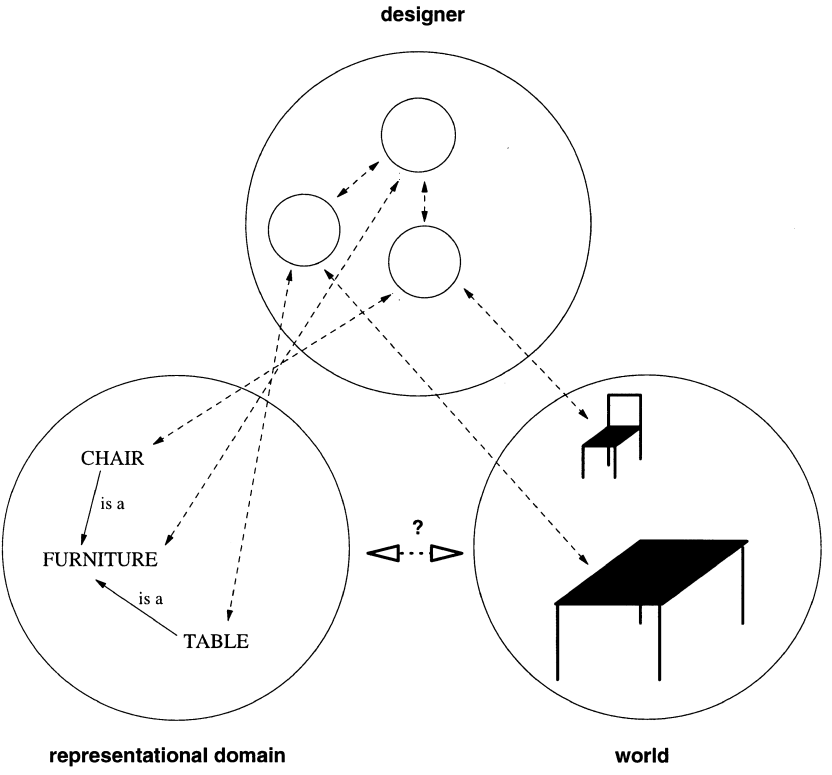

Figure 2. *'What "really" happens in traditional AI representation' (Dorffner 1997). There are direct mappings between objects in the world and the designer's own internal concepts, and between the designer's concepts and their counterparts in the AI system's representational domain. There is, however, no direct, designer-independent, connection between the AI system and the world it is supposed to represent, i.e., the AI system lacks 'first hand semantics' or 'contents for the machine'. Adapted from Dorffner (1997: 101)*

symbolic representations, and should, therefore, be referred to as the problem of *representation grounding* (Chalmers 1992), *concept grounding* (Dorffner and Prem 1993), or the *internalist trap* (Sharkey and Jackson 1994). Searle, however, did not suggest that the idea of intelligent machines would have to be abandoned. In fact, he argued that humans are such machines and that the main reason for the failure of traditional AI was that it is concerned with *computer programs*, but 'has nothing to tell us about *machines*' (Searle 1980), i.e., physical systems causally connected to their environments. That means, instead of accusing AI of being materialistic (for its belief that [man-made] machines, could be intelligent), Searle

actually accused AI of dualism, for its belief that disembodied, i.e., bodyless and body-independent, computer programs could be intelligent.

New AI: Situated and embodied agents

One of the developments of AI and cognitive science in the 1980s was a growing interest in the interaction between agents and their environments. A number of researchers questioned not only the techniques used by traditional AI, but its top-down approach and focus on agent-internal reasoning in general. They suggested a bottom-up approach, also referred to as '*New AI*' or '*Nouvelle AI*', as an alternative to the (purely) computationalist framework of cognitivism. In particular, Brooks (1986b, 1990, 1991a) put forward his *behavior-based robotics* approach and Wilson (1985, 1991) formulated the *animat approach to AI*. These approaches agree that AI should be approached from the bottom up; first and foremost through the study of the interaction between *autonomous agents* and their environments by means of perception and action. For a more detailed review see Ziemke (1998). In this approach, agents equipped with sensors and motors are typically considered *physically grounded* as Brooks explains:

> 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. (Brooks 1990: 6)

These key ideas are also reflected by commitments to 'the two cornerstones of the new approach to Artificial Intelligence, situatedness and embodiment' (Brooks 1991b: 571). The first commitment, to the study of agent-environment interaction rather than representation, is reflected in the notion of *situatedness*: 'The robots are situated in the world — they do not deal with abstract descriptions, but with the here and now of the world directly influencing the behavior of the system' (571). The second commitment was to physical machines, i.e., robotic agents rather than computer programs, as the object of study, as reflected in the notion of *embodiment*: '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' (571).

Thus AI has come (or returned) to an Uexküllian view of semantics, in which signs/representations are viewed as embedded in functional circles along which the interaction of agent and environment is

organized/structured. Or, as T. von Uexküll (1982) put it: '… signs are instructions to operate. They tell the subject (as navigational aids do the seaman) what is to be done, i.e., they give instructions on how to operate' (17). In AI this led to a de-emphasis of representation in the sense of an explicit internal world model mirroring external reality (Brooks 1991a). Instead representations are in the bottom-up approach viewed as *deictic*, i.e., subject-centered, indexical-functional representations (e.g., Agre and Chapman 1987; Brooks 1991b) or 'behavior-generating patterns' (Peschl 1996), i.e., signs that play their role in the functional circle(s) of agent-environment interaction.

Brooks (1986a, 1991a) was also, to our knowledge, the first AI researcher to take inspiration directly from Uexküll's work, in particular the concept of *Merkwelt* or perceptual world. He pointed out that the internal representations in AI programs really were designer-dependent abstractions, based on human introspection, whereas 'as Uexküll and others have pointed out, each animal species, and clearly each robot species with its own distinctly nonhuman sensor suites, will have its own different Merkwelt' (Brooks 1991a: 144). If, for example, in an AI program we had internal representations describing chairs as something one could sit or stand on, that might be an appropriate representation for a human, it would, however, probably be entirely meaningless to a computer or a wheeled robot which could not possibly sit down or climb on top of a chair. Similarly, Uexküll (1982) had pointed out, several decades earlier, that the concept of 'chair' as 'something to sit on' could apply to entirely different objects for a dog than for a human.

Brooks, therefore, approached the study of intelligence through the construction of physical robots, which were embedded in and interacting with their environment by means of a number of so-called *behavioral modules* working in parallel, each of which resembles an Uexküllian functional circle. Each of these behavioral modules is connected to certain receptors from which it receives sensory input (e.g., one module might be connected to sonar sensors, another to a camera, etc.), and each of them, after some internal processing, controls some of the robot's effectors. Further, these modules are connected to each other in some hierarchy, which allows certain modules to subsume the activity of others, hence this type of architecture is referred to as *subsumption architecture* (Brooks 1986b). Thus, a simple robot with the task of approaching light sources while avoiding obstacles, could be controlled by three behavioral modules: one that makes it move forward, a second that can subsume forward motion and make the robot turn when detecting an obstacle with some kind of distance sensors, and a third that can subsume the second and make the robot turn towards the light when detecting a light source

using some kind of light sensor. Thus, using this kind of control architecture, the robot is guided by a combination of taxes working together and in opposition, an idea that can be traced back to the work of Fraenkel and Gunn (1940), who in turn were strongly influenced by Loeb.

A common criticism of Brooks' original subsumption architecture is that it does not allow for learning. Hence, this type of robot, although *operationally autonomous* (cf. Ziemke 1998) in the sense that during runtime it interacts with the environment on its own, i.e., independent of an observer, still remains *heteronomous* in the sense that the largest parts of its functional circles, namely the processing between receptors and effectors, and, thereby, the way it interacts with the environment, is still pre-determined by the designer. A number of researchers have, therefore, pointed out that a necessary element of an artificial agent's autonomy would be the capacity to determine and adapt, at least partly, the mechanisms underlying its behavior (Boden 1994; Steels 1995; Ziemke 1996b, 1998). Different approaches to achieve this are discussed in detail in the next section.

Self-organization of sign processes in artificial organisms

Much research effort during the 1990s has been invested into making robots 'more autonomous' by providing them with the capacity for self-organization. Typically these approaches are based on the use of computational learning techniques to allow agents to adapt the internal parameters of their control mechanisms, and, thus, the functional circles by which they interact with their environment. Different adaptation techniques are described in the next subsection, and it is illustrated how such techniques can allow autonomous agents to adapt their internal sign processes in order to self-organize their sensorimotor interaction, e.g., to determine which environmental stimuli they should respond to, and how. Another subsection then takes this one step further and describes how adaptation techniques have been used to allow groups/populations of agents to self-organize communication among themselves. The differences between conventional mechanisms and artificial organisms are then summarized and discussed in the third subsection.

A 'word of warning': It may seem that much of the following discussion presupposes that robots can have first hand semantics and experience or that they have genuine autonomy, experience, and perception or that the type of learning and evolution we discuss is the same as those in living organisms. That is an incorrect impression, as will be discussed in further detail in the next section (cf. also Sharkey and Ziemke 1998). However,

instead of marking each term with quotes or qualifications such as 'it has been argued that', we have put in this disclaimer so that we can simplify and improve the flow of the discussion.

Robot adaptation

As mentioned above, robot adaptation is typically approached by making the control mechanism, mapping sensory signals to motor commands, adaptive. In particular so-called *artificial neural networks* (ANNs), also referred to as *connectionist networks*, have been used as 'artificial nervous systems' connecting a robot's receptors to its effectors (for collections on this topic see, e.g., Bekey and Goldberg 1993; Brooks et al. 1998; Ziemke and Sharkey 1999). The robots used in this type of research are often mobile robots (see Figure 3 for a typical example), typically receiving sensory input from, for example, infrared (distance) sensors or simple cameras, and controlling the motion of their wheels by motor outputs.

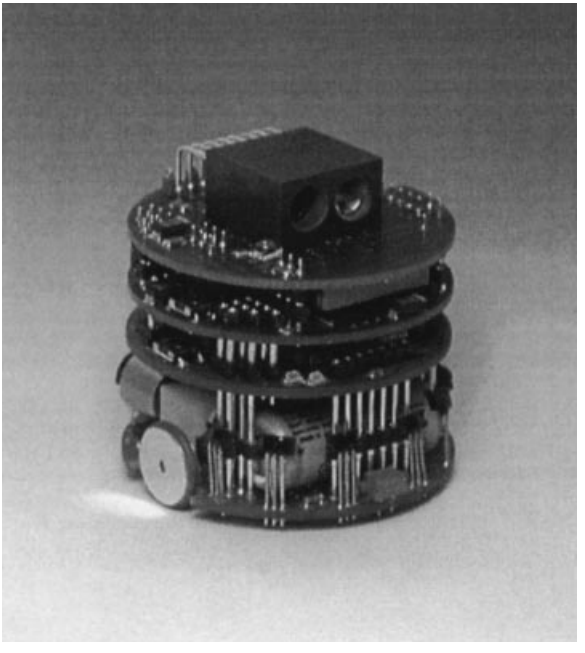

Figure 3. *The Khepera, a wheeled miniature mobile robot commonly used in adaptive robotics research (manufactured by K-Team SA; for details see Mondada et al. 1993). The model shown here is equipped with infrared sensors and a simple camera*

'Artificial nervous systems' for the control of such robots and different learning and evolution techniques for their adaptation will be explained briefly in the following subsections, together with examples of their use in the self-organization of sign processes in artificial organisms.

Artificial neural networks. For the understanding of the argument here it suffices to know that an ANN is a network of a (possibly large) number of simple computational units, typically organized in layers (cf. Figure 4, but note that the number of layers, units, and connection weights can vary greatly). Each unit (or artificial neuron) receives a number of numerical inputs from other units it is connected to, calculates from the weighted sum of the input values its own numerical output value according to some activation function, and passes that value on as input to other neurons, and so on. The feature of ANNs that allows them to learn is the fact that each connection between two units carries a weight, a numerical value itself, that modulates the signal/value sent from one neuron to the other. Hence, by weakening or strengthening of the connection weight, the signal flow between individual neurons can be adapted, and through coordination of the individual weight changes, the network's overall mapping from input to output can be learned.

A number of learning techniques and algorithms have been applied to training neural networks, which vary in the degree of self-organization that they require from the network. During *supervised learning* ANNs are provided with inputs and correct target outputs in every time step, i.e., the network is instructed on which inputs signal to use and which output signals to produce, but how to coordinate the signal flow in between input

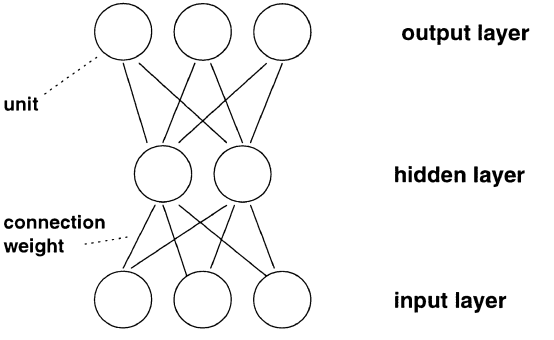

Figure 4. *A typical feed-forward artificial neural network (ANN). Each circle represents a unit (or artificial neuron), and each solid line represents a connection weight between two units. Activation is fed forward only, i.e., from input layer via a hidden layer to the output layer*

and output is up to the network's self-organization. Hence, internal representations (both hidden unit activations and connection weights are commonly interpreted as representations, cf. Sharkey 1991) could be considered to be signs (or their modulators) private to the network and often opaque to outside observers. Thus, unlike traditional AI, connectionists do not promote symbolic representations that mirror a pre-given external reality. Rather, they stress self-organization of an adaptive flow of signals between simple processing units in interaction with an environment, which is compatible with an interactive (Bickhard and Terveen 1995; Bickhard 1998) or experiential (Sharkey 1997) view of representation (see also Clark 1997; Dorffner 1997), and thus offers an alternative approach to the study of cognitive representation and sign use.

Nonetheless, in most connectionist work of the late 1980 and early 1990s, the 'environment' was reduced to input and output values (cf. Clark 1997; Dorffner 1997), i.e., networks were not, like real nervous systems, embedded in the context of an organism and its environment. Thus, although in a technically different fashion, connectionists were, like cognitivists, mainly concerned with explaining cognitive phenomena as separated from organism-world interaction. Hence, they initially focused on modeling isolated cognitive capacities, such as the transformation of English verbs from the present to the past tense (Rumelhart and McClelland 1986) or the pronunciation of text (Sejnowski and Rosenberg 1987), i.e., 'micro-worlds' in Dreyfus' (1979) sense (cf. above discussion). Or in Uexküll's terms: Early connectionism was only concerned with the self-organization of the subject-internal part of the functional circle (where input units might be roughly likened to receptors and output units to effectors). Making the connection between inputs, outputs and internal representations and the actual world they were supposed to represent, was again left to the mind of the observer, similar to the situation illustrated in Figure 2.

Artificial nervous systems. The situation changes fundamentally as soon as ANNs are used as robot controllers, i.e., 'artificial nervous systems' mapping a robot's sensory inputs to motor outputs. Then the network can actually, by means of the robot body (sensors and effectors), interact with the physical objects in its environment, independent of an observer's interpretation or mediation. Hence, it could be argued that its internal signs/representations, now formed in physical interaction with the world they 'represent' or reflect, can be considered physically grounded in the sense explained by Brooks above. Accordingly, the robot controller is in this case part of a complete functional circle (or several circles, as will be discussed below). As an example of this view, imagine a wheeled

robot moving about in a room with boxes lying on the floor and pictures hanging on the wall. The robot might be equipped with infrared sensors as receptors sensitive to the perceptual cues of, for example, the reflectance patterns of solid objects in its environment. Thus, the walls and the boxes on the floor would be part of the robot's own perceptual world (*Merkwelt*), cf. Brooks (1986a). Their meaning to the robot would be that of an 'obstacle', since they limit the robot's motion, assuming the robot has the goal to keep moving while avoiding collisions. Upon detection of the perceptual cue 'solid object at short range' through the distance sensors (receptors) corresponding signs would be transferred to the network's input layer (the robot's 'perceptual organ'). Signs would be transformed and passed on through the internal weights and units of the ANN controlling the robot, and eventually certain signs would reach the output layer (the robot's 'operational organ'), which in turn will transfer signs corresponding to the desired level of activation to the motors controlling the robot's wheels (its effectors). This would make the robot, if trained correctly, move and turn away from the obstacle. Hence, the obstacle or part of it would disappear from the robot's sensor range, such that the receptors would now receive a new perceptual cue, and so on.

The pictures on the wall, on the other hand, would remain 'invisible' to the robot; they are not part of its perceptual world, and they carry no meaning for it. Thus, the robot may be considered to be embedded in its own *Umwelt*, consisting of its perceptual world (*Merkwelt*), consisting of solid objects (or their absence), carrying the meanings 'obstacle' and 'free space' respectively, and its operational world of motor-controlled wheeled motion. The 'inner world' of the robot would be the ANN's internal sign flow and interactive representations, and unlike in the cases of traditional AI programs and Brooks' subsumption architecture, the inner world would here be a self-organized flow of private signs embedded in agent-environment interaction.

Thus learning in ANN robot controllers can be viewed as the creation, adaptation, and/or optimization of functional circles in interaction with the environment. Although the above example illustrated only one such circle, we can, of course, easily imagine several functional circles combined/implemented in a single ANN, e.g., if we additionally equipped the robot with a light and added light sources to the environment, we might have three functional circles: one that makes the robot move forward when encountering 'free space', one that makes it turn/avoid when encountering 'obstacles', and one that makes it approach when detecting the light.

Recurrent ANNs. As long as we are using a feed-forward network, i.e., a network in which activation is only passed in one direction, namely

from input to output units, the mapping from input to output will always be the same (given that the network has already learned and does not modify its connection weights anymore). Hence, the controlled robot will be a 'trivial machine' (cf. T. von Uexküll 1997a), i.e., independent of past or input history, same inputs will always be mapped to same outputs. In semiotic terms this corresponds to a semiosis of information where the input corresponds to the sign, the input-output mapping to the interpretant (or causal rule), and the output to the signified (T. von Uexküll 1997a).

However, if we add internal feedback through recurrent connections to the network, as exemplified in Figure 5, it becomes a 'non-trivial' machine. That means, the mapping from input to output will vary with the network's internal state, and thus the machine, depending on its past, can effectively be a 'different' machine in each time step. An analogy in semiotic terms could be a semiosis of symptomization (cf. T. von Uexküll 1997a) where the interpretant varies and the system's input-output behavior can inform an observer about the current interpretant. For the robot itself this means that it no longer merely reacts to 'external' stimuli, but it interprets signs according to its own internal state. Meeden (1996), for example, trained a toycar-like robot using a recurrent controller network (of the type illustrated in Figure 5a; originally introduced by

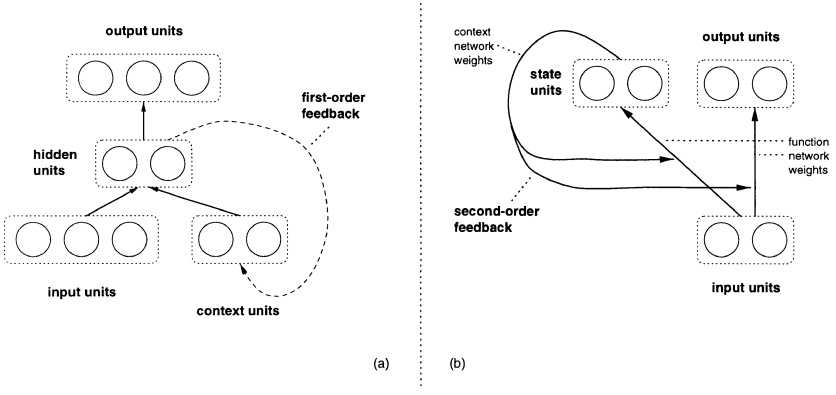

Figure 5. *Recurrent artificial neural networks (RANNs), using (a) first-order feedback, and (b) second-order feedback. Solid arrows indicate that each unit in the first layer of units (layers are surrounded by dotted lines) is connected to each unit in the second layer. The dashed arrow in (a) represents a copy-back connection. That means, hidden unit activation values at time step* t *are fed-back and re-used as extra-inputs at time step* (t + 1). *In (b) the function network weights, i.e., the connection weights between and input and output/state units, embodying the (current) sensorimotor mapping, can be adapted dynamically via a feedback loop (through the context network weights)*

Elman 1990) to periodically approach and avoid a light source while avoiding other obstacles. Information on whether to avoid or to seek the light at a particular point in time was not available/accessible to the robot from the environment (in some of the experimental setups). Instead the control network developed in the learning process an internal dynamic, i.e., a way of utilizing its own feedback signs, that allowed it to form a purely internal hidden unit representation of its current goal. That means, here the functional circles connecting robot (subject) and light source (object), and thus the light cue's meaning, do actually vary with time, not completely unlike the varying level of hunger effects the meaning a piece of food has for an animal.

The recurrent networks discussed so far utilize first-order feedback. That means, as illustrated in Figure 5a, previous activation values are used as extra inputs to certain neurons (typically at the input layer) in later time steps (typically the next). Hence, the network output is in each time step computed as a result of the current input and the context of an internal state (referred to as 'context units' in Figure 5a). A second-order networks, on the other hand, is exemplified in Figure 5b. Here second-order (i.e., multiplicative) feedback (through state units and context network weights), is used to dynamically adapt the connection weights between input and output units (the function network weights). Thus, the mapping from sensory input to motor output can effectively be adapted from time step to time step, depending on an internal state (referred to as 'state units' in Figure 5b). For a detailed description of different variations of this type of network and examples of its use for robot adaptation see Ziemke (1996a, 1996c, 1997, 1999a).

Hence, in this type of controller the sensorimotor mapping, and thus the controlled agent's *behavioral disposition* (or interpretant), dynamically changes with the agent's internal state. Ziemke (1999a), for example, documents experiments in which a Khepera robot, controlled by a second-order network, encounters identical objects inside and outside a circle, but has to exhibit two very different responses to the exact same stimuli (approach inside the circle, and avoidance outside). The problem is that the robot cannot sense whether or not it currently is inside or outside the circle, but only senses the boundary line while passing it on its way in or out. The robot learns/evolves to solve the problem by dynamically adapting its behavioral disposition (interpretant), i.e., its behavioral/motor biases and the way it responds to stimuli from the objects it encounters. This means that, depending on its current behavioral disposition, the robot attributes different meanings to the object stimuli, such that the exact same stimulus can adopt very different *functional tones* (cf. Uexküll 1957) in different contexts.

Reinforcement learning. For complex tasks robots are typically not trained using supervised learning techniques. This has two reasons: (a) In order to allow for a maximum of robot autonomy, it is often desirable to reduce designer intervention to a minimum of feedback/instruction, and (b) it is often not even possible to provide a robot with an exact target output in every time step, for much the same reason why it is impossible to tell a child learning to ride a bike how exactly to move its legs, arms, and body at every point in time. For such tasks, the robot, much like the child, simply has to figure out for itself how exactly to solve a problem, i.e., how to organize and adapt its sign processes in interaction with the environment. Hence, robots are often trained using *reinforcement learning* or *evolutionary adaptation* techniques.

During reinforcement learning (RL), an agent is provided only with occasional feedback, typically in terms of positive and negative reinforcement, e.g., in the case of Meeden's robot when hitting an obstacle ('bad') or achieving a light goal ('good'). From this feedback the agent can adapt its behavior to the environment in such a way as to maximize its positive reinforcement and minimize its negative reinforcement. Reinforcement, in this context, is simply defined as a stimulus which increases the probability of the response upon which it is contingent.

Walter (1951) was the first to use RL techniques for the training of robots. By grafting the Conditioned Reflex Analogue (CORA), a learning box, onto *Machina speculatrix* (cf. discussion above), he created *Machina docilis*, the easy learner. *M. docilis* had built-in phototaxis, i.e., a light elicited a movement response towards it which he referred to as 'an unconditioned reflex of attraction'. When a light was repeatedly paired with the blowing of a whistle, *M. docilis* became attracted to the sound of the whistle and exhibited a phonotaxic response. In a separate series of experiments, Walter repeatedly paired the sound of the whistle with obstacle avoidance and thus trained the robot to 'avoid' the sound of the whistle. He also demonstrated extinction of conditioned pairings by presenting the conditioned stimulus repeatedly without pairing it with the unconditioned stimulus. There was also a slower decay of the conditioned response if it was not used for some time. Walter's experiments show how a simple learning mechanism can extend the behavior of a robot by bringing its reflexes under the control of substituted environmental effects.

Evolutionary adaptation. The use of evolutionary techniques is an approach to 'push' the designer even further 'out of the learning loop' and aims to let robots learn from the interaction with their environment with a minimum of human intervention (cf. Nolfi 1998). Evolutionary methods are abstractly based on the Darwinian theory of natural selection.

Thus, feedback is no longer instructive as in reinforcement and supervised learning, but only evaluative. Typically, a population of individuals (e.g., robot controllers) is evolved over a large number of generations, in each of which certain individuals are selected according to some fitness function, and 'reproduced' into the next generation, using recombinations and slight mutations mimicking natural reproduction. Due to the selective pressure the average fitness in the population is likely to increase over generations, although the individuals typically do not learn during their 'lifetime'. The very idea of evolving robots was well illustrated by Braitenberg (1984) who likened evolution to the following scenario: There are a number of robots driving about on a table top. At approximately the same rate that robots fall off the table, others are picked up randomly from the table, one at a time, and copied. Due to errors in the copying process, the original and the copy might differ slightly. Both are put back onto the table. Since the fittest robots, those who stay on the table longest, are most likely to be selected for 'reproduction' the overall fitness of the robot population is likely to increase in the course of the 'evolutionary' process.

A concrete example of *evolutionary robotics* research is the work of Husbands et al. (1998) who evolved RANN robot controllers for a target discrimination task, which required a mobile robot, equipped with a camera, to approach a white paper triangle mounted on the wall, but to avoid rectangles. In these experiments both the network topology and the visual morphology (or receptive field), i.e., which parts/pixels of the camera image the controller network would use as inputs, were subject to the evolutionary process. The analysis of the experimental runs showed that structurally simple control networks with complex internal feedback dynamics evolved which made use of low bandwidth sensing (often only two pixels of visual input were used) to distinguish between the relevant environmental stimuli. Thus, in these experiments both the internal flow of signals and use of feedback, as well as the 'external' sign use, i.e., which environmental stimuli to interpret as signs of what, are the result of an artificial evolutionary process. The evolved sign processes are difficult to analyze and understand in detail, due to the fact that they are *private* to the robot and in many cases radically differ from the solutions the human experimenters would have designed. Husbands et al. point out that this 'is a reminder of the fact that evolutionary processes often find ways of satisfying the fitness criteria that go against our intuitions as to how the problem should be ''solved'' ', (Husbands et al. 1998: 206).

The influence of the human designer can be reduced even further using *co-evolutionary methods*. Nolfi and Floreano (1998), for example,

co-evolved two RANN-controlled robots to exhibit predator- and prey-behavior. The 'predator', a Khepera robot equipped with an extra camera (cf. Figure 3) which allowed it to observe the prey from a distance, had to catch (make physical contact with) the 'prey', another Khepera robot, equipped only with short-range infrared sensors but also with the potential to move faster than the 'predator'. By simply evolving the two 'species' with time-to-contact as a fitness and selection criterion, quite elaborate pursuit- and escape-strategies evolved in the respective robots. The predator species, for example, in some cases developed a dynamics that allowed it to observe and interpret the prey's current behavior as a symptom of its current behavioral disposition, and thus of its behavior in the immediate future, such that it would only 'strike' when it had a realistic chance of catching the prey 'off guard'.

The examples discussed so far have only been concerned with the evolution/adaptation of artificial nervous systems. Recently, however, researchers have begun to apply evolutionary methods also to the construction of physical structures and robot morphologies (in simulation) (e.g., Funes and Pollack 1997; Lund et al. 1997), in some cases in co-evolution with controllers (Cliff and Miller 1996; Lund and Miglino 1998). Cliff and Miller (1996), for example, simulated the co-evolution of 'eyes' (optical sensors) and 'brains' (ANN controllers) of simple robotic agents which pursued and evaded each other in a two-dimensional plane. The co-evolution of both body and 'brain' of artificial organisms aims to overcome what Funes and Pollack called the 'chicken and egg' problem of the approach: 'Learning to control a complex body is dominated by inductive biases specific to its sensors and effectors, while building a body which is controllable is conditioned on the pre-existence of a brain' (1997: 358). For a detailed discussion of the epistemological implications of robotic devices which evolve/construct their own hardware see Cariani (1992).

In summary, we have seen a number of examples of artificial organisms self-organizing (a) their internal usage of signs (their 'inner world'), in the form of ANN connection weights, (b) the way they respond to stimuli from the environment, and in some cases (c) the way they dynamically self-adapt their behavioral disposition, i.e., the way they make use of internal sign usage to adapt their response to 'external' stimuli. Thus, in many of these examples, it is left up to a process of self-organization, to determine which of the objects in the environment become carriers of meaning, and what exactly their meaning is to the agent. The next section will take this one step further, and illustrate how adaptive techniques have been used by populations of agents to facilitate the self-organization of communication between them.

Self-organized communication in autonomous agents

Traditional AI research initially focused on endowing computers with human-level cognitive capacities, of which natural language communication was by many considered to be of particular relevance. Alan Turing (1950), a key figure in the development of the very idea of AI, in fact, suggested the (later) so-called *Turing test* as a criterion for machine intelligence. In this test a machine would have to carry on a natural language conversation on arbitrary every day topics with a human judge for a certain period of time, via some teletype-terminal so the judge could not see whether he is communicating with a machine or a human being. If, after that time, the judge could not reliably identify the machine as a machine, it would, according to Turing, have to be considered to possess human-level intelligence. This test was considered a valid criterion of intelligence by most AI researchers at least until the 1980s, and many AI systems simulating human communication were built. Most famous among them was perhaps Weizenbaum's (1965) ELIZA system, which simulated a human psychiatrist.

From the arguments of Dreyfus, Searle, and others (cf. above), however, it became clear that, of course, in these conversations the AI system performed purely syntactic transformations of the symbols it was fed, without even a clue of their actual meaning. That means the AI system processed a language ('natural' to the human observer) without actually understanding it. On some reflection this is not too surprising, after all what could a conversation about the objects of human experience (like tables, chairs, etc.) possibly mean to a computer system completely lacking this type of experience? In Uexküll's and Brooks' terms, even if a computer program had a perceptual world, it would be very unlikely to contain, for example, chairs since certainly it could not sit on them or make any other meaningful use of them.

The study of communication has, therefore, been addressed in a radically different way in AI and ALife research since about the mid-1990s. Now communication is studied from the bottom up, i.e., using autonomous agents that can actually 'experience' and interact with their environment. Moreover, artifacts are no longer expected to learn human language, but their own language, i.e., a language that is about 'the world as it appears to them' and that helps them to communicate with other agents (no longer humans) in order to better cope with that world.

In the spirit of the bottom-up approach, these communication systems must be developed by the robots themselves and not designed and programmed in by an external observer. They must also be grounded in the sensori-motor experiences

of the robot as opposed to being disembodied, with the input given by a human experimenter and the output again interpreted by the human observer. (Steels and Vogt 1997: 474)

Cangelosi and Parisi (1998), for example, have in computer simulations studied the evolution of a 'language' in a population of ALife agents that 'live' in a simulated world containing edible and poisonous mushrooms, of which they have to find the former but avoid the latter in order to ensure survival. The agents were controlled by ANNs which received as input 'sensory' information about mushrooms nearby and produced as output 'motor' commands that controlled the agent's motion. Additionally, each agent could output communication signals, which other agents could receive as additional input. The scenario was set up such that agents would profit from communicating, i.e., every agent approaching a mushroom required the help of another agent telling it whether the mushroom was edible or not. The results showed that after 1,000 generations of artificial evolution the agents had indeed evolved a simple 'language' of signals that allowed them to communicate about the world they 'lived' in, i.e., the approach and avoidance of the mushrooms they encountered.

Experiments on the development of 'language' and 'meaning' in groups of robotic agents through 'adaptive language games' have been carried out by Steels (1998; see also Steels and Vogt l997; Steels and Kaplan 1999). In the experimental setup used by Steels and Vogt (1997), a number of mobile robots moved around in a physical environment of limited size, containing some additional objects. The robots acquired a common 'vocabulary' of word-meaning pairs (where the meaning of a word is taken to be the sensory feature set it is associated with) through 'adaptive language games', which work roughly as follows. Whenever two robots meet they first perform a simple 'dance' in the course of which they turn 360 degrees and scan the view of their environment. They agree on some sensory feature set, e.g., a box nearby, and both focus on it. Then both robots check if they already have a 'word' for the object/feature set they see. If only one of them has, it tells the other, which now learns the new word. If neither of them has a word for the object, they 'make one up', and both learn it. If both already know different words for the object, one of them forgets the old word and learns a new one from the other robot. After that the robots begin roaming the environment separately again. Since there are several robots, a common 'language' develops and eventually spreads to the whole population through the accumulative transfer, creation, and adaptation of a common vocabulary as a result of the development and interaction of individual lexica of word-meaning

pairs in the course of the one-to-one language games performed by the robots. For a discussion of the semiotic dynamics resulting from this kind of experiment, e.g., the emergence and dampening of synonymy and polysemy; see also Steels and Kaplan (1999).

Thus, in both these examples autonomous agents are not 'forced' to learn a human language they could not, due to their radically different physiology, possibly understand. Instead they develop, in a process of self-organization, their own language from the interaction with their environment and other agents, i.e., a language that is specific to their 'species', in the sense that it is based on their *own* experience and serves their *own* purposes, and thus is not necessarily interpretable to human observers (cf. Dorffner and Prem 1993; Prem 1995).

It could, however, be argued (cf. Prem 1998), that this type of approach to the evolution/development of language is misguided in that it is typically based on the old symbol/representation grounding idea of hooking independently existing external objects to abstract internal labels/signs (cf. Figure 1). An example is the above work of Steels and Vogt in which the sensory feature set that a word is associated with is taken to be its meaning. In Jakob von Uexküll's view of signs, however, as Thure von Uexküll (1982) put it: 'Signs are instructions to operate' which 'tell the subject … what is to be done', i.e., signs derive their meaning from the role they play in the functional circles of the interaction between a subject and its object(s). Communication should, therefore, perhaps first and foremost be addressed as giving agents the possibility to influence each others' behavior. That means, they should be able to communicate signals that help them to interact or coordinate their behavior instead of learning a vocabulary without actual functional value for the interaction between agent and environment (cf. Ziemke 1999b), as in the above case of Steels and Vogt, where the agents never actually use those object labels for anything more than just the labeling of objects.

How artificial organisms differ from conventional mechanisms

We have now seen a number of examples of autonomous agents and their self-organization. Together these examples illustrate that artificial organisms, although certainly mechanisms in the technical sense, in a number of points radically differ from the type of mechanism that Uexküll discussed, and, in fact, exhibit some of the properties that he ascribed to organisms alone. This subsection summarizes the differences between artificial organisms and other mechanisms. The following section will then complement this one by taking an in-depth look at the differences between

artificial and living organisms, and the implications for their respective autonomy and capacity for semiosis.

Firstly, the use of 'artificial nervous systems' in combination with computational learning techniques allows autonomous agents to adapt to their environment. In particular, due to their use of memory the behavioral disposition of autonomous agents varies over time. Thus, although they do not 'grow' in the physical sense, they do adapt to their environment, such that they do, in fact, have a 'historical basis of reaction' (cf. the arguments of Driesch and Uexküll discussed above). Self-organized artificial organisms thus no longer react in a purely physical or mechanical manner to causal impulses. Instead their reaction carries a 'subjective' quality, in the sense that the way they react is not determined by built-in rules (alone), but is specific to them and their history of 'experience' and self-organization.

Secondly, and closely related to the previous point, artificial organisms are clearly involved in sign processes, and they 'make use' of signs 'themselves', unlike the mechanisms Uexküll discussed. Furthermore, unlike computer programs which are to some degree also capable of semiosis (cf. Andersen et al. 1997 and the discussion in the introductory section), the sign processes of artificial organisms are typically (a) not (fully) determined by their human designers, (b) independent of interpretation through external observers (at least at the operational level), and (c) in many cases not even interpretable to humans through a close look at the internal processes (despite the fact that these are much easier to observe than in the case of a living organism). Much of the sign usage of such systems is, therefore, due to their self-organization, indeed *private* and specific to them. Artificial organisms, therefore, have been argued to have a certain degree of *epistemic autonomy* (Prem 1997; cf. also Bickhard 1998), i.e., like living organisms they are 'on their own' in their interaction with their environment.

Thirdly, the use of self-organization, especially evolutionary techniques, *does* nowadays (to some degree) allow the construction of robot controllers, and to some degree even robot bodies (in simulation), following *centrifugal principles*. In the context of robot controllers, Nolfi formulated the concept as 'adaptation is more powerful than decomposition and integration' (Nolfi 1997a, 1997b). Here controllers are not, as in Brooks' subsumption architecture or conventional robot design, broken down into behavioral or functional modules by a designer, but the task decomposition is the result of a process of adaptation, which distributes behavioral competences over subsystems in a modular architecture. Similarly, as mentioned above, in some of the first author's work (e.g., Ziemke 1996a, 1999a), the control of a robot is broken into a number

of functional circles in a process of dynamic adaptation and differentiation. In these cases the control mechanism is not constructed along centripetal principles, i.e., not broken down into sub-tasks or -competences by a designer to be integrated later, but instead constructed making use of what might be called *centrifugal task decomposition*. That means, a single control mechanism breaks itself down into a number of sub-mechanisms in a process of adaptation and differentiation. Similar principles have even been applied to the co-evolution of physical structures and robot morphologies with controllers (e.g., Cliff and Miller 1996; Lund and Miglino 1998). Here robot body and controller are no longer treated as isolated elements to be constructed separately, but instead they are co-evolved in an integrated fashion as the result of the evolution of a single artificial genotype. The use of centrifugal principles (although not under that name) has during the 1990s become a 'hot topic' in ALife research, and there are various approaches to the combination of evolution, development, and learning in the self-organization of artificial organisms. Another example is the work of Vaario and Ohsuga (1997) on 'growing intelligence' which integrates processes of development, learning, natural selection, and genetic changes in simulated artificial organisms.

The role of the living body

Having illustrated the principles of artificial organisms and their self-organization and having outlined the differences between such systems and conventional mechanisms in the previous section, we will now turn to the differences between artificial and living organisms. The next subsection presents a brief comparison between Uexküll's theory and the work of Maturana and Varela on autopoiesis and the biology of cognition. The implications of the lack of a living body for the autonomy of artificial organisms and their sign processes are then considered in the second subsection.

Uexküll versus Maturana and Varela

As discussed above, the (re-) turn to artificial organisms in AI research can be seen as a rejection of the purely computationalist framework of traditional cognitive science. Instead, work on ALife and autonomous robots has to some degree taken inspiration from the work of Humberto Maturana and Francisco Varela, who have since the late 1960s developed

their theories on the biology of cognition and autopoiesis (e.g., Maturana 1969; Varela 1979; Maturana and Varela 1980, 1987) which has more recently also lead to the formulation of an enactive cognitive science (Varela et al. 1991). To summarize their work goes beyond the scope of this article. It is, however, worth pointing out the relation to the unfortunately less known, but closely related and much earlier work of Jakob von Uexküll in a number of points, in particular since Maturana and Varela apparently themselves were not aware of Uexküll's work.

Maturana and Varela's work is strictly opposed to the cognitivist framework of traditional cognitive science, and instead is aimed at understanding the biological basis of cognition. They propose a way of 'seeing cognition not as a representation of the world "out there", but rather as an ongoing bringing forth of a world through the process of living itself' (Maturana and Varela 1987: 11). This somewhat unconventional use of the term 'cognition' may be clarified by Bourgine and Varela's (1992) characterization of the *cognitive self* (similar to Uexküll's notion of 'subject') as the 'specific mode of coherence, which is embedded in the organism':

> … the cognitive self is the manner in which the organism, through its own self-produced activity, becomes a distinct entity in space, though always coupled to its corresponding environment from which it remains nevertheless distinct. A distinct coherent self which, by the very same process of constituting itself, configures an external world of perception and action. (Bourgine and Varela 1992: xiii)

Similar to Uexküll's emphasis of the subjective nature of living organisms, Maturana and Varela (1987) point out that 'all cognitive experience involves the knower in a personal way, rooted in his biological structure'. In particular they characterize living organisms, as well as the cells they consist of, as *autopoietic unities*, i.e., self-producing and -maintaining systems, and like Uexküll they 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). (Maturana and Varela 1987: 51)

Maturana and Varela distinguish between the organization of a system and its structure. The *organization*, similar to 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: 47). Living systems, for example, 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. This is 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: 47). Thus the structure of an autopoietic system is the concrete realization of the actual components (all of their properties) 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 as some of the mechanistic theories would suggest. Machines made by humans, including cars and robots, 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. Again, it is worth pointing out that, despite differences in terminology, Maturana and Varela's distinction between autopoietic and allopoietic machines, is very similar to Uexküll's (1928) earlier discussed 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.

The two-way fit between organism and environment is what Maturana and Varela refer to as *structural congruence* between them, which is the

result of their *structural coupling*:

> Ontogeny is the history of structural change in a unity without loss of organization in that unity. This ongoing structural change occurs in the unity from moment to moment, either as a change triggered by interactions coming from the environment in which it exists or as a result of its internal dynamics. As regards its continuous interactions with the environment, the … unity classifies them and sees them in accordance with its structure at every instant. That structure, in turn, continuously changes because of its internal dynamics. …
>
> In these interactions, the structure of the environment only *triggers* structural changes in the autopoietic unities (it does not specify or direct them), and vice versa for the environment. The result will be a history of mutual congruent structural changes as long as the autopoietic unity and its containing environment do not disintegrate: there will be a *structural coupling*. (Maturana and Varela 1987: 74)

Moreover, similar to Uexküll's (1928) view of autonomous cellular unities (*Zellautonome*) as the basic components of multicellular organisms, Maturana and Varela refer to the former as 'first-order autopoietic unities' and to the latter as 'second-order autopoietic unities', and they characterize their integration/solidarity as follows:

> … in the dynamism of this close cellular aggregation in a life cycle, 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. (Maturana and Varela 1987: 79)

Finally, it should be mentioned that, although they are compatible in many aspects, there are, of course, differences between the two theoretical frameworks compared here. For example, Uexküll's outright rejection of evolutionary theories in general, and the work of Darwin in particular, is a position that in its strictness now, more than fifty years later, appears untenable (cf. Emmeche 1990; Hoffmeyer 1996). Maturana and Varela's work, although also skeptical towards neo-Darwinism and its overly simplistic view of 'natural selection', is certainly more in agreement with modern evolutionary theory. In fact, the view of evolution as 'natural drift' is an important element of their theory of the biology of cognition (for details see Maturana and Varela 1987; Varela et al. 1991). A common criticism of Maturana and Varela's theory of autopoiesis, on the other hand, is its disregard for such concepts as representation[10] and information (cf. Emmeche 1990). Hence, in this aspect many cognitive scientists, and certainly many researchers in semiotics, will probably prefer the theoretical framework of Uexküll whose theories emphasize the central role of sign processes in all aspects of life.

On the differences between artificial and living organisms

Having discussed the differences between artificial organisms and conventional mechanisms above, this section will examine what exactly the (remaining) differences between living and artificial organisms are, and what semiotic relevance these differences have. In the following discussion concepts from both the theories of Uexküll as well as Maturana and Varela will be used. We do so because we believe that the two theoretical frameworks, concerning the issue at hand, are sufficiently compatible, and, in fact, enrich each other.

As discussed in the previous sections, modern AI research on the interaction between artificial organisms and their environments has, unlike the still predominant computer metaphor, certainly taken a lot of inspiration from biology. Nevertheless, modern autonomous robotics and ALife research, due to its interest in (intelligent) behavior and its focus on observability, often sidesteps much of the proximal details, i.e., the actual biology, and goes directly for the behavior. Thus, robots are enabled to interact with their environment, such that a distal description of the behavior, at the right level, can be compared with the description of some living systems' behavior at the same level of description (e.g., 'obstacle avoidance'). Thus, the Turing test has been replaced by a behavioral test. If, for example, a robot avoids obstacles or follows a cricket's calling song (cf. Lund et al. 1998) 'just like a real animal', then the internal processes of artificial and living organism are taken to be equivalent, at least possibly. This, however, is just the observer's interpretation of the robot's behavior. On some reflection, nobody would suggest that the robot following the male cricket's calling song actually does so in order to mate. Uexküll (1982: 36) pointed out that the 'life-task of the animal … consists of utilizing the meaning carriers and the meaning-factors, respectively, according to their particular building-plan'. It might be argued that the calling song 'carries meaning' for both female cricket and robot, in the sense that they both utilize the signal in order to move towards its source. Ultimately, however, we have to acknowledge that this behavior is meaningful only for the cricket (or its species), since it contributes to the fulfilling of its 'life-task'. In the robot's case this behavior is only meaningful to the observer, simply because the robot has no 'life-task' independent of observation, and its phonotaxis is not at all part of a coherent whole of agent and environment (cf. also Sharkey and Ziemke 1998, 2000).

The robot's relation to its environment is very different from the living organism's, i.e., the 'embodiment' and 'situatedness' of natural organisms are far more deeply rooted than those of their artificial counterparts

(cf. Ziemke 1999b). A robot might have self-organized its control system, possibly even its physical structure to some degree, in interaction with its environment, and thus have acquired a certain degree of 'epistemic autonomy' (Prem 1997; Cariani 1992). This self-organization, however, 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 lie in the robot's components themselves and success or failure of the process is not 'judged' by them either. 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. Neither do they suddenly have an intrinsic 'life-task', even in an abstract sense; the 'task' still is in the head of the observer. The 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). T. von Uexküll (1997a) has pointed out that living organisms are autopoietic systems (cf. previous subsection), which selectively assimilate parts of their environment and get rid of parts they do not need anymore. According to T. von Uexküll, selection and assimilation of the required elements can be described as sign processes, whose interpretants correspond to the living systems' biological needs. The criterion for the correctness of the interpretation described by the sign process is the successful assimilation. Robots, however, do not assimilate anything from their environment, and, as mentioned above, they have no intrinsic needs that the self-organizing process would have to fulfill to remain 'viable'. 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 research on artificial organisms, we believe, is that, despite claims to the contrary and despite the emphasis of 'embodiment', many researchers are still devoted to the *computationalist/functionalist* view of *medium independence*, i.e., the idea that the 'characteristics of life and mind are independent of their respective material substances' (Emmeche 1992: 471). Much research effort is spent on control mechanisms, or 'artificial nervous systems', and how to achieve certain behaviors in robots through self-organization of these control mechanisms. However, to compare a robot's 'artificial nervous system' to an animal's nervous system, because they exhibit 'the same behavior', implies that the relation between behavior and (artificial) nervous system is actually independent of

the controlled body. In other terms, it implies that the operation of the nervous system is computational and largely independent of the body it is carried out in, i.e., the body is reduced to the computational control system's sensorimotor interface to the environment. Maturana and Varela, however, have argued (again, similar to Uexküll 1928; cf. also Hoffmeyer 1996), that in living organisms body and nervous system 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. (Maturana and Varela 1987: 34)

Similarly, T. von Uexküll et al. (1993, 1997), in their discussion of *endosemiosis*, 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 earlier discussed neural counterworld (cf. 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).

In summary, it can be said that, despite all biological inspiration, artificial organisms are still radically different from their living counterparts. 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 of living organisms. Mostly, this is due to the fact that artificial organisms are composed of mechanical parts and control programs. The autonomy and subjectivity of living systems, on the other hand, emerges from the interaction of their components, i.e., autonomous cellular unities (*Zellautonome*). Meaningful interaction

between these first-order unities, and between the resulting second-order unity and its environment, is a result of their structural congruence, as pointed out by Uexküll, as well as Maturana and Varela. Thus, autonomy is a property of a living organism's organization right from its beginning as an autonomous 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 is 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 'create' 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 to adapt its structure in order to solve a task defined by its designer (cf. also Sharkey and Ziemke 2000, where we discuss the relation to the case of Clever Hans [Pfungst 1911]). Thus, the lack of autonomy makes the idea of true 'first hand semantics' or 'content for the machine' in today's robotic systems highly questionable.

Summary and conclusions

The aim of this article has been to discuss the relation between Jakob von Uexküll's work and contemporary research in AI and cognitive science. In particular, we have used his theory of meaning to evaluate the semiotic relevance of recent research in adaptive robotics and ALife.

The article started off by discussing Uexküll's and Loeb's views of the differences between organisms and mechanisms, as well as early attempts at putting mechanistic theories to the test through the construction of artificial organisms. Then AI's attempts to create a new type of mechanism, which should have some of the mental and/or behavioral capacities of living organisms, was discussed. It was noted that, after three decades of focusing on disembodied computer programs, AI research returned to its cybernetic roots, and now again much research is devoted to the interaction between agents and their environments.

The autonomous agents approach to AI and ALife has incorporated influences from a number of theories. From the work of Loeb and others

the view that organisms are more or less guided by the environment through taxes/tropisms has found its way into robotics, and has become very influential. From cognitivism many researchers, perhaps without much reflection, have adopted the general idea that the nervous system carries out computation, mapping sensory inputs to motor outputs. However, the bottom-up approach distances itself strongly from the cognitivist correspondence view of representation as a 'mirror' of a pre-given world and instead focuses on interactive representations as behavior-guiding structures (Bickhard and Terveen 1995; Peschl 1996; Dorffner 1997; Ziemke 1999a). This is much in line with Uexküll's view of signs as embedded in the functional circles of agent-environment interaction. Moreover, Uexküll influenced Brooks' (1986a, 1991a) argument that, like any living organism, an autonomous agent would have to have its own 'subjective' view of the world.

Further, it was then discussed how 'artificial nervous systems' in combination with computational learning techniques are used in the attempt to make artificial organisms (more) autonomous by enabling them to self-organize their sign processes. Several examples illustrated how such techniques allow robots to find their own way of organizing their functional circles i.e., their internal use of signs and their response to stimuli from the environment. It was further pointed out that the use of self-organization and memory does indeed make artificial organisms a unique type of mechanism that might be of further semiotic interest.

The previous section then first examined the relation between Uexküll's theory and Maturana and Varela's work on embodied cognition and its biological basis. It can be noted that the two theoretical frameworks, both developed against the 'mainstream', are largely compatible, although (unfortunately) developed independently. Moreover, the differences between living and artificial organisms were examined in further detail. It was pointed out that, despite all biological inspiration and self-organization, today's so-called 'autonomous' agents are actually far from possessing the autonomy of living systems. This is mostly due to the fact that artificial organisms are composed of mechanical parts and control programs. Living organisms, on the other hand, derive their autonomy and 'subjectivity' from their cellular autonomous unities' integration and structural congruence with the environment, as pointed out by Uexküll as well as Maturana and Varela. Together with the fact that artificial organisms simply lack an intrinsic 'life task', this strongly questions the idea of 'first hand semantics' or 'content for the machine' in today's robotic systems.

However, it has been shown that the AI/ALife community strives to minimize human intervention in the design of artificial organisms and

actively investigates alternative, more 'life-like' ways of 'constructing' such systems. So far self-organization through adaptation in interaction with the environment has mostly been applied to control systems, but it has also been discussed that researchers are beginning to apply similar approaches to the integrated self-construction of robot bodies and nervous systems. For future work along these lines a greater awareness of Jakob von Uexküll's work would be important, since it could help to avoid the pitfalls of 'new' overly mechanistic theories. We believe that his theories will prove to be of great value to researchers in robotics, ALife, and embodied cognition in their endeavor to gain further understanding of the meaningful embedding of living organisms in their worlds, as well as the possibilities and limitations of their artificial counterparts.[11]

Notes

1. To avoid confusion between Jakob von Uexküll and his son Thure, we will throughout the article refer to both authors by first and last name, or to the former as 'Uexküll' and to the latter as 'T. von Uexküll'.
2. Sebeok, for example, writes (in personal communication cited by T. von Uexküll 1982) that 'the criterial feature of living entities, and of machines programmed by humans, is semiosis'.
3. See Uexküll's figure of the functional cycle in the beginning of this issue.
4. Langthaler (1992), with reference to T. von Uexküll, points out that Uexküll's view, although often associated with vitalism, should really be considered a 'third position', combining elements of both mechanism and vitalism. In a similar vein Emmeche (this issue) argues that Uexküll's theory, as well as modern biosemiotics in general, should be considered a kind of qualitative organicism.
5. We here use the translation given by T. von Uexküll et al. (1993), who translate the original German term 'Zeichen' as 'sign', rather than 'token' as in the earlier translation given in Uexküll (1985).
6. All our translations from German sources have been carried out by the first author (who is a native speaker).
7. A number of similar examples, built in the first half of the twentieth century, has been discussed by Slukin (1954).
8. 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.
9. See also Hoffmeyer (1996: 47) who argues (not specifically directed at AI though) that 'mental ''aboutness'' — human intentionality — grew out of a bodily ''aboutness'' (i.e., the behavior necessary for assuring reproduction and survival)' and points out that we 'cannot escape the fact that our minds remain embodied'.
10. See, however, also Varela et al.'s (1991) more recent formulation of an enactive cognitive science, which is to a large extent compatible with an interactive view of representation.
11. The authors would like to thank Claus Emmeche and Kalevi Kull for helpful comments on an earlier version of this article. Tom Ziemke is supported by a grant (1507/97) from the Knowledge Foundation, Stockholm.

References

Agre, Philip E. and Chapman, David (1987). Pengi: An implementation of a theory of activity. In *Proceedings of the Sixth National Conference on Artificial Intelligence AAAI-87*, 268–272. Los Angeles: Morgan Kaufmann.

Andersen, Peter B.; Hasle, Per; and Brandt, Per A. (1997). Machine semiosis. In *Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, Roland Posner, Klaus Robering, and Thomas A. Sebeok (eds.), 548–571. Berlin: Walter de Gruyter.

Bekey, George and Goldberg, K. Y. (eds.) (1993). *Neural Networks in Robotics*. Boston: Kluwer.

Bickhard, Mark H. (1998). Robots and representations. In *From Animals to Animats 5: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, Rolf Pfeifer (ed.), 58–63. Cambridge, MA: MIT Press.

Bickhard, Mark H. and Terveen, L. (1995). *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution*. New York: Elsevier.

Boden, Margaret A. (1996). *The Philosophy of Artificial Life*. Oxford: Oxford University Press.

Bourgine, Paul and Varela, Francisco J. (1992). Toward a practice of autonomous systems. In *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, F. J. Varela and P. Bourgine (eds.), xi–xvii. Cambridge, MA: MIT Press.

Braitenberg, Valentino (1984). *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: MIT Press.

Brooks, Rodney A. (1986a). *Achieving Artificial Intelligence through Building Robots*. (Technical Report Memo 899.) Cambridge, MA: MIT AI Lab.

—(1986b). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation* 2, 14–23.

—(1990). Elephants don't play chess. *Robotics and Autonomous Systems* 6 (1/2), 3–15.

—(1991a). Intelligence without representation. *Artificial Intelligence* 47, 139–159.

—(1991b). Intelligence without reason. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, 569–595. San Mateo: Morgan Kaufmann.

Brooks, Rodney A.; Grossberg, S.; and Optican, L. (eds.) (1998). *Neural Control and Robotics: Biology and Technology*. *Neural Networks* 11 (7/8). [Special issue.]

Cangelosi, Angelo and Parisi, Domenico (1998). The emergence of a 'language' in an evolving population of neural networks. *Connection Science* 10 (2), 83–97.

Cariani, Peter (1992). Some epistemological implications of devices which construct their own sensors and effectors. In *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, F. J. Varela and P. Bourgine (eds.), 484–493. Cambridge, MA: MIT Press.

Chalmers, David J. (1992). Subsymbolic computation and the Chinese room. In *The Symbolic and Connectionist Paradigms: Closing the Gap*, J. Dinsmore (ed.), 25–48. Hillsdale, NJ: Lawrence Erlbaum.

Clancey, William J. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*. New York: Cambridge University Press.

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

Cliff, Dave and Miller, G. F. (1996). Co-evolution of pursuit and evasion II: Simulation methods and results. In *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, P. Maes, M. Mataric, Jean-Arcady Meyer, J. B. Pollack, and S. W. Wilson (eds.), 506–515. Cambridge, MA: MIT Press.

Craik, Kenneth J. W. (1943). *The Nature of Explanation*. Cambridge: Cambridge University Press.

Darwin, Charles (1859). *The Origin of Species*. London: John Murray.

De Candolle, A. P. (1832). In Fraenkel and Gunn 1940.

Dorffner, Georg (1997). Radical connectionism — A neural bottom-up approach to AI. In *Neural Networks and a New Artificial Intelligence*, G. Dorffner (ed.), 93–132. London: International Thomson Computer Press.

Dorffner, Georg and Prem, Erich (1993). Connectionism, symbol grounding, autonomous agents. In *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, 144–148. Hillsdale, NJ: Lawrence Erlbaum.

Dreyfus, Hubert (1979). *What Computer Can't Do*, revised edition. New York: Harper and Row.

Driesch, Hans (1931). *Das Wesen des Organismus*. Leipzig: Barth.

Elman, Jeffrey (1990). Finding structure in time. *Cognitive Science* 14, 179–211.

Emmeche, Claus (1990). Kognition og omverden — om Jakob von Uexküll og hans bidrag til kognitionsforskningen. *Almen Semiotik* 2, 52–67.

—(1992). Life as an abstract phenomenon: Is artificial life possible? In *Toward a Practice of Autonomous Systems — Proceedings of the First European Conference on Artificial Life*, F. J. Varela and P. Bourgine (eds.), 466–474. Cambridge, MA: MIT Press.

—(2001). Does a robot have an Umwelt? *Semiotica*, this issue.

Fraenkel, G. and Gunn, D. L. (1940). *The Orientation of Animals: Kineses, Taxes and Compass Reactions*. Oxford: Clarendon Press.

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

Funes, Pablo and Pollack, Jordan B. (1997). Computer evolution of buildable objects. In *Proceedings of the Fourth European Conference on Artificial Life*, Phil Husbands and Inman Harvey (eds.), 358–367. Cambridge, MA: MIT Press.

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

Hoffmeyer, Jesper (1996). *Signs of Meaning in the Universe*. Bloomington: Indiana University Press.

Husbands, Phil; Smith, Tom; Jakobi, Nick; and O'Shea, Michael (1998). Better living through chemistry: Evolving gasnets for robot control. *Connection Science* 10 (3/4), 185–210.

James, William (1961 [1892]). *Psychology: A Briefer Course*. New York: Harper.

Johnson-Laird, Philip N. (1989). Mental models. In *Foundations of Cognitive Science*, M. I. Posner (ed.), 469–493. Cambridge, MA: MIT Press.

Knight, T. A. (1806). In Fraenkel and Gunn 1940.

Langthaler, Rudolf (1992). *Organismus und Umwelt — Die biologische Umweltlehre im Spiegel traditioneller Naturphilosophie*. Hildesheim: Georg Olms Verlag.

Lloyd, Dan E. (1989). *Simple Minds*. Cambridge, MA: MIT Press.

Loeb, Jacques (1918). *Forced Movements, Tropisms, and Animal Conduct*. Philadelphia, PA: Lippincott.

Lund, Henrik Hautop; Hallam, John; and Lee, Wei-Po (1997). Evolving robot morphology. In *Proceedings of the IEEE Fourth International Conference on Evolutionary Computation*, 197–202. New York: IEEE Press.

Lund, Henrik Hautop and Miglino, Orazio (1998). Evolving and breeding robots. In *Proceedings of the First European Workshop on Evolutionary Robotics*, 192–211. Berlin: Springer.

Lund, Henrik Hautop; Webb, Barbara; and Hallam, John (1998). Physical and temporal scaling considerations in a robot model of cricket calling song preference. *Artificial Life* 4, 95–107.

Maturana, Humberto R. (1969). The neurophysiology of cognition. In *Cognition: A Multiple View*, Paul Garvin (ed.), 3–24. New York: Spartan Books.

Maturana, Humberto R. and Varela, Francisco J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: D. Reidel.

—(1987). *The Tree of Knowledge — The Biological Roots of Human Understanding*. Boston: Shambhala. [Second, revised edition 1992.]

Meeden, Lisa (1996). An incremental approach to developing intelligent neural network controllers for robots. *IEEE Transactions on Systems, Man, and Cybernetics: Part B, Cybernetics* 26 (3), 474–485.

Mondada, Francesco; Franzi, E.; and Ienne, P. (1994). Mobile robot miniaturisation: A tool for investigating in control algorithms. In *Experimental Robotics III: Proceedings of the 3rd International Symposium on Experimental Robotics, Kyoto, Oct 1993*, T. Yoshikawa and F. Miyazaki (eds.), 501–513. London: Springer Verlag.

Morris, Charles W. (1946). *Signs, Language, and Behavior*. Englewood Cliffs, NJ: Prentice Hall.

Müller, Johannes (1840). *Handbuch der Physiologie des Menschen*, Band 2. Koblenz: J. Hölscher.

Neisser, Ulric (1967). *Cognitive Psychology*. New York: Appelton.

Nolfi, Stefano (1997a). Evolving non-trivial behavior on autonomous robots: Adaptation is more powerful than decomposition and integration. In *Evolutionary Robotics '97: From Intelligent Robots to Artificial Life*, Takashi Gomi (ed.), 21–48. Ottawa: AAI Books.

—(1997b). Using emergent modularity to develop control systems for mobile robots. *Adaptive Behavior* 5 (3/4), 343–363.

—(1998). Evolutionary robotics: Exploiting the full power of self-organization. *Connection Science* 10 (3/4), 167–183.

Nolfi, Stefano and Floreano, Dario (1998). Co-evolving predator and prey robots: Do 'arm races' arise in artificial evolution? *Artificial Life* 4 (4), 311–335.

Palmer, Stephen E. (1978). Fundamental aspects of cognitive representation. In *Cognition and Categorization*, E. Rosch and B. B. Lloyd (eds.), 259–303. Hillsdale, NJ: Erlbaum.

Peschl, Markus F. (1996). The representational relation between environmental structures and neural systems: Autonomy and environmental dependency in neural knowledge representation. *Nonlinear Dynamics, Psychology and Life Sciences* 1 (3), 99–121.

Pfeffer, Wilhelm F. P. (1883). In Fraenkel and Gunn 1940.

Pfungst, Oskar (1911). *Clever Hans (The Horse of Mr. von Osten): A Contribution to Experimental Animal and Human Psychology*. New York: Henry Holt.

Prem, Erich (1995). Understanding complex systems: What can the speaking lion tell us? In *The Biology and Technology of Autonomous Agents* (=NATO ASI Series F 144), L. Steels (ed.), 459–474. Berlin: Springer.

—(1996). *Motivation, Emotion and the Role of Functional Circuits in Autonomous Agent Design Methodology* (=Technical Report 96-04). Vienna: Austrian Research Institute for Artificial Intelligence.

—(1997). Epistemic autonomy in models of living systems. In *Fourth European Conference on Artificial Life (ECAL97)*, Phil Husbands and Inman Harvey (eds.), 2–9. Cambridge, MA: MIT Press.

—(1998). Semiosis in embodied autonomous systems. In *Proceedings of the IEEE International Symposium on Intelligent Control*, 724–729. Piscataway, NJ: IEEE.

Rumelhart, David E. and McClelland, Jay L. (1986). On learning the past tense of English verbs. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*, D. E. Rumelhart and J. L. McClelland (eds.), 216–271. Cambridge, MA: MIT Press.

Rylatt, Mark; Czarnecki, Chris A.; and Routen, Tom W. (1998). Beyond physical grounding and naive time: Investigations into short-term memory. In *From Animals to Animats 5: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, Rolf Pfeifer (ed.), 22–31. Cambridge, MA: MIT Press.

Searle, John (1980). Minds, brains and programs. *Behavioral and Brain Sciences* 3, 417–457.

Sejnowski, Terrence and Rosenberg, C. (1987). Parallel networks that learn to pronounce English text. *Complex Systems* 1, 145–168.

Sharkey, Noel E. (1991). Connectionist representation techniques. *Artificial Intelligence Review* 5, 143–167.

—(1997). Neural networks for coordination and control: The portability of experiential representations. *Robotics and Autonomous Systems* 22 (3/4), 345–359.

Sharkey, Noel E. and Jackson, Stuart A. (1994). Three horns of the representational trilemma. In *Artificial Intelligence and Neural Networks: Steps towards Principled Integration*, Vasant Honavar and Leonard Uhr (eds.), 155–189. Boston: Academic Press.

Sharkey, Noel E. and Ziemke, Tom (1998). A consideration of the biological and psychological foundations of autonomous robotics. *Connection Science* 10 (3/4), 361–391.

—(2000). Life, mind and robots — The ins and outs of embodiment. In *Hybrid Neural Systems*, Stefan Wermter and Ron Sun (eds.), 313–332. Heidelberg: Springer.

Sherrington, Charles S. (1906). *The Integrative Action of the Nervous System*. New York: C. Scribner's Sons.

Slukin, W (1954). *Minds and Machines*. Middlesex, UK: Penguin.

Steels, Luc (1995). When are robots intelligent autonomous agents? *Robotics and Autonomous Systems* 15, 3–9.

—(1998). The origin of syntax in visually grounded robotic agents. *Artificial Intelligence* 103, 133–156.

Steels, Luc and Kaplan, Frederic (1999). Situated grounded word semantics. In *IJCAI-99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, T. Dean (ed.), 862–867. San Francisco: Morgan Kauffmann.

Steels, Luc and Vogt, Paul (1997). Grounding adaptive language games in robotic agents. In *Fourth European Conference on Artificial Life*, Phil Husbands and Inman Harvey (eds.), 474–482. Cambridge, MA: MIT Press.

Strasburger, Eduard (1868). In Fraenkel and Gunn 1940.

Turing, Alan (1950). Computing machinery and intelligence. *Mind* 59, 433–460.

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

—(1973 [1928]). *Theoretische Biologie*. Berlin: Springer Verlag. [All page numbers refer to the 1973 edition, Frankfurt: Suhrkamp.]

—(1957). A stroll through the worlds of animals and men: A picture book of invisible worlds. In *Instinctive Behavior: The Development of a Modern Concept*, C. H. Schiller (ed.), 5–80. New York: International Universities Press. [Also in *Semiotica* 89 (4), 319–391.]

—(1982). The theory of meaning. *Semiotica* 42 (1), 25–82.

—(1985). Environment [Umwelt] and inner world of animals. In *The Foundations of Comparative Ethology*, Gordon M. Burghardt (ed.), 222–245. New York: Van Nostrand Reinhold. [Partial translation of Uexküll (1909).]

Uexküll, Thure von (1982). Introduction: Meaning and science in Jakob von Uexküll's concept of biology. *Semiotica* 42 (1), 1–24.

—(l992). Introduction: The sign theory of Jakob von Uexküll. *Semiotica* 89 (4), 279–315.

—(1997a). Biosemiose. In *Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, Roland Posner, Klaus Robering, and Thomas A. Sebeok (eds.), 447–457. Berlin: Walter de Gruyter.

—(1997b). Jakob von Uexkülls Umweltlehre. In *Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, Roland Posner, Klaus Robering, and Thomas A. Sebeok (eds.), 2183–2191. Berlin: Walter de Gruyter.

Uexküll, Thure von; Geigges, Werner; and Herrmann, Jörg M. (1993). Endosemiosis. *Semiotica* 96 (1/2), 5–51.

—(1997). Endosemiose. In *Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, Roland Posner, Klaus Robering, and Thomas A. Sebeok (eds.), 464–487. Berlin: Walter de Gruyter.

Vaario, Jari and Ohsuga, S. (1997). On growing intelligence. In *Neural Networks and a New Artificial Intelligence*, G. Dorffner (ed.), 93–132. London: International Thomson Computer Press.

Varela, Francisco J. (1979). *Principles of Biological Autonomy*. New York: Elsevier North Holland.

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

Walter, William Grey (1950). An imitation of life. *Scientific American* 182 (5), 42–45.

—(1951). A machine that learns. *Scientific American* 184 (8), 60–63.

—(1953). *The Living Brain*. New York: W. W. Norton.

Weizenbaum, J. (1965). Eliza — A computer program for the study of natural language communication between man and machine. *Communications of the ACM* 9, 36–45.

Wilson, Stewart W. (1985). Knowledge growth in an artificial animal. In *Proceedings of an International Conference on Genetic Algorithms and Their Applications*, J. Grefenstette (ed.), 16–23. Hillsdale, NJ: Lawrence Erlbaum.

—(1991). The animat path to AI. In *From Animals to Animats: Proceedings of The First International Conference on Simulation of Adaptive Behavior*, J-A. Meyer and Stewart Wilson (eds.), 15–21. Cambridge, MA: MIT Press.

Ziemke, Tom (1996a). Towards adaptive behaviour system integration using connectionist infinite state automata. In *From Animals to Animats 4 – Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, P. Maes, M. Mataric, J-A. Meyer, J. B. Pollack, and S. W. Wilson (eds.), 145–154. Cambridge, MA: MIT Press.

—(1996b). Towards autonomous robot control via self-adapting recurrent networks. In *Artificial Neural Networks: ICANN 96*, C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff (eds.), 611–616. Berlin: Springer.

—(1996c). Towards adaptive perception in autonomous robots using second-order recurrent networks. In *Proceedings of the First Euromicro Workshop on Advanced Mobile Robots (EUROBOT '96)*, 89–98. Los Alamitos: IEEE Computer Society Press.

—(1997). The 'environmental puppeteer' revisited: A connectionist perspective on 'autonomy'. In *Proceedings of the 6th European Workshop on Learning Robots (EWLR-6)*, Andreas Birk and John Demiris (eds.), 100–110. Brighton: Artificial Intelligence Laboratory.

—(1998). Adaptive behavior in autonomous agents. *Presence* 7 (6), 564–587.

—(1999a). Remembering how to behave: Recurrent neural networks for adaptive robot behavior. In *Recurrent Neural Networks: Design and Applications*, Larry Medsker and L. C. Jain (eds.), 355–389. New York: CRC Press.

—(1999b). Rethinking grounding. In *Understanding Representation in the Cognitive Sciences: Does Representation Need Reality?*, Alexander Riegler, Markus Peschl, and Astrid von Stein (eds.), 177–190. New York: Plenum Press.

Ziemke, Tom and Sharkey, Noel E. (eds.) (1998). *Biorobotics*. Special issue of *Connection Science* 10 (3/4).

—(eds.) (1999). *Artificial Neural Networks for Robot Learning*. Special issue of *Autonomous Robots* 7 (1).

Tom Ziemke (b. 1969) is Assistant Professor of Computer Science at the University of Skövde, Sweden ⟨tom@ida.his.se⟩. His research interests include neuro-robotics, situated action and theories of interactive representation, embodied cognition and their biological basis. His publications include 'Adaptive behavior in autonomous agents' (1998), 'A consideration of the biological and psychological foundations of autonomous robotics' (with N. E. Sharkey, 1998), 'Learning and unlearning mechanisms in animats and animals' (with T. Savage, 2000), and 'The construction of reality in the robot' (2001).

Noel E. Sharkey (b. 1948) is Professor of Computer Science at the University of Sheffield in the United Kingdom ⟨noel@dcs.shef.ac.uk⟩. His research interests include adaptive robotics, training neural net controllers, self-learning robots (evolutionary and neural computing), bio-robotics, sensor-fusion, embodied cognition, and developmental intelligence and engineering applications. His major publications include 'Fundamental issues in connectionist natural language processing' (1996), 'Separating learning and representation' (with A. J. C. Sharkey, 1996), 'The new wave in robot learning' (1997), and 'Learning from innate behaviors: A quantitative evaluation of neural network controllers' (1998).