Does a robot have an Umwelt? Reflections on the qualitative biosemiotics of Jakob von Uexküll

CLAUS EMMECHE

Introduction

How does the Umwelt concept of Jakob von Uexküll fit into current discussions within theoretical biology, philosophy of biology, biosemiotics, and Artificial Life, particularly the research on 'autonomous systems' and robots? To investigate this question, the approach here is not historical Uexküll scholarship exposing the original core of philosophical ideas that provided an important background for the original conception of the Umwelt in the writings of Jakob von Uexküll (some of which seem incompatible with a modern evolutionist perspective); rather, I will show that some aspects of his thoughts are still interesting and provide inspiration in contemporary biology, cognitive science, and other fields. Therefore, I will also draw upon his son Thure von Uexküll's reflections in his further development of the Umwelt theory, which is not anti-evolutionary (his father's approach was anti-Darwinian, which is not the same as anti-evolutionary though often interpreted as such).

Specifically, I will investigate the plausibility of three theses: (1) The Umwelt theory of Jakob von Uexküll, even though his theoretical biology was often characterized as being thoroughly vitalist, can in the context of contemporary science, more adequately be interpreted as a branch of qualitative organicism in theoretical biology. *Qualitative organicism* is a position which claims, first, a kind of middle road position, that is, on the one hand, there are no mysterious or non-material vital powers in organisms (non-vitalism), but on the other hand, the characteristic properties of living beings cannot be fully accounted for by physics and chemistry because these properties are nonreducible emergent properties (emergentism); second, that some of these emergent properties have an experiential, phenomenal, or subjective character which plays a major role in the dynamics of the living system. Modern biosemiotics (inspired by C. S. Peirce and Jakob von Uexküll, instituted by

Semiotica 134–1/4 (2001), 653–693 0037–1998/01/0134–0653

Thomas A. Sebeok) is a kind of qualitative organicism. (2) This position sheds light on recent discussions in cognitive science, artificial life, and robotics about the nature of representation and cognition — indeed genuine semiotic questions as they deal with the role of information and signs for any system that has the property of being 'animal-like,' that is, systems that move by themselves and seem to be guided by a kind of entelechy or, in modern but shallow terms, a behavioral program. (3) Particularly, qualitative organicism allows us to approach the question of whether a robot can have an Umwelt in the sense that Jakob von Uexküll used the term (a subjectively experienced phenomenal world). The eventuality of a positive answer to this question, i.e., a claim that a robot indeed *can* have an Umwelt, seems counterintuitive to the extent that a robot may be seen as — to use a bewildering word — an incarnation of the mechanical and reductionist world picture to which Jakob von Uexküll was so strongly opposed. But certain ideas and concepts may sometimes lead us to unexpected consequences, which threaten our cherished metaphysical assumptions, and we should try to face such questions with an open mind.

Asking this third question, we must also inquire if that is the same as asking 'Can the robot have a mind'? If so, the Umwelt is just another word for the concept of mind, and the theory of Jakob von Uexküll would not contribute to solve our question. But this is clearly not the case. Though one might think, that if one has a very broad concept of mind, e.g., motivated by biosemiotics and the philosophy of Peirce, then the *Umwelt* of animals and the *mind* of animals might have the same extension. However, mind as such is not co-extensive with Umwelten, at least for the semiotic notion of mind one finds in Peirce (Santaella Braga 1994). The mind is a broader notion than the Umwelt, so, for instance, there can be a lot of activity in a living organism which is of a mental, or semiotic, character, but which does not figure as a part of the animal's experienced phenomenal world. Clearly, the two concepts mean different things and neither have the same intension nor extension. I do not know whether Peirce would have ascribed mind-like properties to robots, but it appears to be the case that he would — both biological organisms and robots with sensors and effectors could in principle embody the same logical or semiotic principles (cf. Burks 1975).

The main route below is through the following points. After a short introduction to the Umwelt concept of Jakob von Uexküll (see also other articles in this issue), his theory will be situated in the tradition of qualitative organicism in biology, to be introduced. Let me emphasize that the Umwelt theory may be interpreted in other ways (such as being strictly vitalist), so what I intend is not a critical exposition of

J. von Uexküll's own version but a reconstruction of the theory more in line with contemporary theoretical biology. The next step is a historical overview of research in robotics and autonomous systems, a scientific field that has already attracted the attention of semioticians (cf. Meystel 1998) and which is deeply inspired by biological considerations. Along with that overview and in a separate section hereafter, the question whether a robot can have an Umwelt will be finally decided (hopefully). The perspectives of a closer approachment of theoretical biology, semiotics, autonomous systems research, and cognitive science are discussed.

The Umwelt concept and present day biology

Umwelt: Not environment, not mind

The Umwelt may be defined as the phenomenal aspect of the parts of the environment of a subject (an animal organism), that is, the parts that it selects with its species-specific sense organs according to its organization and its biological needs (J. von Uexküll 1940; T. von Uexküll 1982a, 1989). In that sense, the subject is the constructor of its own Umwelt, as everything in it is labelled with the perceptual cues and effector cues of the subject. Thus, one must at least distinguish between these concepts: (1) the *habitat* of the organism as 'objectively' (or externally) described by a human scientific observer; (2) the *niche* of the organism in the traditional ecological sense as the species' ecological function within the ecosystem, (3) the *Umwelt* as the experienced self-world of the organism.[1]

The Umwelt notion deeply influenced Konrad Lorenz in his development of ethology, but it never really became established within ethology or general biology and was subsequently forgotten for a long period. This may in part be due to the dominance of Darwinian thinking in biology and the fact that Jakob von Uexküll very early had become a convinced anti-Darwinist and was also subsequently associated with the vitalist opposition against mechanicism in biology (see Harrington 1996 for additional biographical information). Lorenz, of course, was a Darwinist. As Richards (1987: 530) remarks,

> Despite Lorenz's adamantine commitment to ultra-Darwinism, his instinct-theory bore the sign of an openly anti-Darwinian thinker — Jakob von Uexküll, a Drieschian vitalist. From Uexküll, an independent scholar of spousal means, Lorenz adapted the notion of a 'functional system' (*Funktionskreis*). According

to Uexküll's theory, a functional or interactive system constituted the relation between an animal, with its special organs and needs, and its own experienced world (*die Umwelt*), the lived reality of which corresponded to the animal's sensory abilities and requirements. Lorenz transformed Uexküll's conception of the functional system into that of the 'innate releasing schemata' (*angeborenen Auslöse-Schemata*). This innate releasing mechanism (IRM), as he also termed it, was the receptor correlate in the animal that responded with a particular pattern of behavior to specific elicitory cues in the environment.

This passage hints at the historical fact that the focus on the very phenomenal aspect of the Umwelt in the subsequent development of the mainstream study of animal behavior was quickly toned down and almost completely disappeared, probably because of influences upon ethology from such intellectual movements as positivism, behaviorism, and, in biology, neo-Darwinism and mechanicism.

However, with the development in the second half of this century of zoosemiotics, biosemiotics, and psychosomatic medicine, the Umwelt notion came increasingly into use again. This short quotation from a paper on 'Endosemiosis' by Thure von Uexküll, Wernes Geigges, and Jörg M. Hermann suffices to restate the core of the concept: 'Jakob von Uexküll coined the term *Umwelt* (''subjective universe'', ''significant surround'', ''phenomenal world'', or ''self-world'', as opposed to *Umgebung* — ''environment''; […])' (T. von Uexküll et al. 1993: 6). In a note here the authors elaborate that an Umwelt is the subjective world of what is meaningful impingement for the living being in terms of its information processing equipment, sign system, and codes. They continue to note that animals are wrapped in networks of sign processes which protect them by transposing the environment into its subjective meaning, which is accessible only to the encoding subject.

Two features of the Umwelt notion are important in this context: (1) As noted, the system's Umwelt cannot be equated with the system's mind. By whatever means one characterizes mind, its activity is more encompassing than what specifically is experienced by the system as its world. For instance in humans, our Umwelt becomes conscious by means of intentional perception, cognition, and language while an ocean of subconscious or non-conscious processes are active parts of the mind. (2) An organism has only primary access to its own Umwelt, and only humans (and some rather clever 'mind-reading' animals, such as certain predators interpreting the mind of their prey) may by inferences have indirect access to the Umwelt of other species. However, this 'indirect access' is never the same thing as the real Umwelt of the species in question — e.g., our scientific understanding of the sonar system of a bat gives us an indirect and functional picture of the bat's Umwelt,

but we cannot enter into that Umwelt itself; all we have is a model in our (linguistic, cognitive, and perceptual) Umwelt of the bat's Umwelt. Science attempts to build a model-based 'view from nowhere' (Nagel 1986), but can only do so mediated by our species-specific Umwelt, our subjective point-of-view from which we collectively construct a shared human sphere of public knowledge.

Qualitative organicism

An often seen misinterpretation is the construal of philosophy of twentieth-century biology as a fight between vitalism and mechanicism that finally was won by mechanicism. This construal overlooks the fact that the most influential position turned out to be organicist (even though popular science after the advent and triumphs of molecular biology told a different story to the public). The 'resolution of the debate' between vitalism and mechanicism was not a mechanist stance, but a sort of historical compromise in the form of what I here call *mainstream organicism* (exemplified by the writings of such well-known biologists as J. Needham, P. Weiss, C. H. Waddington, J. Woodger, E. Mayr, R. C. Lewontin, R. Levins, and S. J. Gould) functioning more or less tacitly as a background philosophy of biology. For some of those who have devoted much intellectual energy to fight reductionist thinking in biology (which is indeed common but also, most often, merely programmatic) this interpretation may sound surprising, but one should distinguish between ill-founded spontaneous talk about organisms as being merely mechanical aggregates of molecules, and the real conceptual structure and scientific practice within domains like evolutionary or molecular biology where reduction to chemistry or physics is never really the issue. In science, metaphysical attachments and scientific research may be connected, but often only loosely so: It is quite possible for adherents of metaphysical vitalism, organicism, and mechanicism to work together in the same laboratory progressing in substantiating the same paradigm, abstaining from philosophical conflicts or restricting themselves to an instrumentalist discourse. Scientists often have a very pragmatic stance towards foundations, an attitude which is characteristic also of mainstream organicism. It is, nevertheless, possible to explicate that position. Organicism takes the complexity and physical uniqueness of the organism as a sign of the distinctiveness of biology as a natural science *sui generis*.[2] This position has several historical roots; one precursor is the emergentist movement at the beginning of the twentieth century, especially in Britain.[3] This middle

road, although here often framed within a naturalist evolutionary perspective, was anticipated by Kant's more critical (non-naturalist) notion of a living organism.[4] According to Kant, we cannot dispense with a heuristic principle of purposefulness when we consider an organism, that is to say, '*An organized product of nature is one in which every part is reciprocally purpose [end] and means*. In it nothing is vain, without purpose, or to be ascribed to a blind mechanism of nature' (Kant 1790 [1951]: 222). However, within mainstream organicism this teleology is interpreted as a more or less 'mechanical' teleonomy being the result of the forces of blind variation and natural selection, plus eventually some additional 'order for free' or physical self-organization. Mainstream organicism as a position is thus non-vitalist, ontologically non-reductionist (allowing for methodological reduction) and emergentist. What are studied as emergent properties are common material structures and processes within several levels of living systems (developmental systems, evolution, self-organizing properties, etc.), all of which are treated in the usual way as objects with no intrinsic experiential properties. For instance, in behavioral studies, the ethologists are not allowed to make use of subjectivist or anthropocentric language describing animal behavior.

In contrast, *qualitative organicism* represents a more 'colored' view of living beings; it emphasizes not only the ontological reality of biological higher level properties or entities (such as systems of self-reproducing organisms being parts of the species' historical lineages) but also the existence of phenomenological or qualitative aspects of at least some higher level properties. When sensing light or colors, an organism is not merely performing a detection of some external signals which then get processed internally (described in terms of neurochemistry or 'information processing' or whatever); something additional is going on (at least if we want the full story), namely the organism's own experience of the light, and this experience is seen as something very real. Even though it has a subjective mode of existence, it is an objectively real phenomenon. (In recent philosophy of mind, Searle [1992] is one of the few to emphasize the ontological reality of subjective experience; however, he is, most of the time, only talking about human experience.) As a scientific position qualitative organicism is concerned with qualities which are not only of the famous category of 'primary' qualities (roughly corresponding to the scientifically measurable quanta), including shape, magnitude, and number; but also concerned with the 'secondary' qualities of color, taste, sound, feeling, etc.[5] One should not equate qualitative organicism or mainstream organicism with coherent stances, theories, or paradigms; though for both options one can find

representatives in recent theoretical biology.[6] Some authors may not be consistent, some may only implicitly express either idea; the important thing is to recognize that, in fact, two different conceptions of life and biosemiosis are at stake.

It is obvious that the Umwelt notion is of central importance to the development of a coherent theory of the qualitative experiential world of the organism, a task present day biology must face, instead of continuing to ignore a huge phenomenal realm of the living world — the experiential world of animal appetites, desires, feelings, sensations, etc.[7] For such a task, theoretical inspiration can be found in the fields of semiotics as well as artificial life and autonomous systems research. The experiential Umwelt is rooted in the material and semiotic body of the organism, which again is situated in a specific part of the habitat depending on its (Eltonian) niche. An actual theory of the Umwelt must not posit any vitalist spiritual or occult hidden powers to 'explain' the emergence of the Umwelten in evolution; however, it must acknowledge the richness and reality of the phenomena of organismic sensing, acting, and perceiving. The implication of such an adventure could be important not only to biology, but as well to semiotics (to ground the sign notion in nature), to philosophy of mind (to overcome dualism and solve the problems of non-reductive supervenience physicalism), and to a general understanding of the relation between the human and other species. Could we create an artificial 'organism' with an Umwelt more alien to us than that of a chimp or a fruit fly?

Autonomous systems: A brief history

It is often suggested that many devices that are able to serve specific human purposes could be more convenient and useful if they could be 'autonomous agents', that is, not only be computational input-output devices, but move around as cybernetic systems by their own motor modules guided by sensors, making decisions, having the capacity of acting more or less intelligently given only partial information, learning by their mistakes, adapting to heterogeneous and changing environments, and having a sort of life of their own. No doubt such devices could also cause a lot of harm (but here I'll disregard concerns about technology assessment and ethical implications). Various research programs have been launched and are running for the study and design of what is called 'autonomous systems', 'situated agents', 'distributed AI systems', and 'multi-agent systems' — not only for the purpose of

designing 'useful' agents, but often mainly to investigate what it really is for a system to be autonomous and have some sort of agency.

This field of research, here denoted ASR (autonomous systems research), is continuous with classical Artificial Intelligence research (AI) in several aspects, especially in its implicit *structuralism*: The aim is not so much the scientific study of natural forms of the phenomenon (intelligent behavior) as its general and more abstract processual structure, obviously to see if other instances of its structure could be designed artificially to solve specific problems. In the case of classical AI, dating back to the 1950s and 1960s and still existing today, the purpose was not so much the scientific study of *human* intelligence, which more became the focus of cognitive science (CS), as it was (and still is) the creation of a cluster of theories of *possible intelligent* systems that can be implemented in physical instances of Turing machines. With the invention of Artificial Life (AL) as a research program in the mid-1980s (the first conference was in 1987), the theoretical purpose was the study of 'life as it could be', to extend, so to speak, the base set of examples provided by traditional 'carbon-chauvinist' biology which was blamed for having dealt empirically only with a single class of living systems — those that accidentally happened to have evolved on Earth.[8] Likewise, the study of autonomous systems (which partly builds upon and is continuous with topics in AI, CS, and AL) is not so much focused on the causal structure of naturally realized autonomous systems (microbes, plants, animals, humans) as it is focused on the structure of any conceivable system that can possibly realize autonomous behavior.

This structuralism may be viewed as important for creative design and engineering of new systems types, a necessary liberation from the focus on empirical investigation of naturally realized systems. However, within AI, and certainly also within ASR, it has created epistemological confusion in relation to *the Pygmalion Syndrome* (cf. Emmeche 1994a: 63, 134–155). This is the fallacy of taking an artificially created model not only to be *representing* reality but to *be* just another instance of reality. If a computational AL system, such as Tom Ray's TIERRA (in Langton et al. [eds.] 1992), is taken not simply to model some abstract aspects of evolution by natural selection, but to be an instance of life, one has committed the Pygmalion fallacy. There has been an extensive debate whether this is really a fallacy or a part and parcel of the 'strong' AL program.[9] Similarly, one could claim that the devices created within the field of ASR are either just more or less interesting *models* of 'real' living autonomous organisms (where the artificial systems are not intrinsically autonomous, because the property of

autonomy is ascribed to them in their function as a model), or that they are simply cybernetic machines that certainly may behave *as if* they were autonomous, but where this autonomy is either too simple to catch the intended property of the real thing, or simply of another category of behavior. The concept of an Umwelt is seldom used in this discussion,[10] although a real understanding of the Umwelt concept may have deep implications for the possibility of 'strong AL' (the idea of not simply simulating or imitating life processes but creating *genuine* life *de novo*, from scratch so to say, by artificial means).

The point here is not to say that research in autonomous systems, animats, and robotics has failed because these systems will never become 'truly autonomous', or that they are biologically unrealistic, useless, wacky, or similar allegations — this may indeed be so for some or all specific systems created until now, but one should acknowledge a rich and varied field of research and technological development that can inspire and invigorate not only the coming industry of 'intelligent design' but also a lot of scientific investigations. It is a very open question what will be achieved by this research in the future, and the point here is to articulate questions, rather than give definitive answers. A crucial question is whether such systems can have an Umwelt, and if so, how would it look and how could we know, and if not, why not? Approaching these questions involves investigation of the metaphysical ideas and presuppositions of this research. What kinds of systems are really autonomous; is it the ones with an intrinsic relation between Umwelt and autonomy?

A note on terminology. The term 'autonomous' in the literature of ASR is used in a variety of ways, most often with informal meanings. Autonomous means in ordinary language a person, a region, or a state that is self-governing; independent; subject to its own laws.[11] A connotation is freedom, such as freedom of the will. Such connotations depend, of course, on what kind of system is viewed as being autonomous.[12] The term autonomous derives from the Greek word auto-, or *autos* meaning self, the same, and *nomos* meaning law, i.e., self-governing, self-steering, spontaneous, opposed to heteronomous meaning externally controlled. In the biological theory of Maturana and Varela (1980), the term was given a specific meaning, viz., the condition of subordinating all changes to the maintenance of organization and 'the self-asserting capacity of living systems to maintain their identity through the active compensations of deformations' (1980: 135).[13] However, within ASR, what counts as an 'autonomous agent' would often be classified as being a non-autonomous (heteropoietic) system by the criteria given by their theory of autopoiesis.

Cybernetics, robotics, classical AI: Some historical forerunners

The idea of autonomous systems originates both in pre-scientific ideas of what constitutes adaptive, intelligent task-solving behavior in man, animals, and machines, and in the early attempts to model and construct systems with seemingly goal-directed behavior during the early period of cybernetics, information theory, and related disciplines (systems theory, operation theory, and general engineering science). Also other fields became crucial for this development later on, such as automatic reasoning, pattern recognition, 'intelligent' data bases, expert systems, other AI techniques, and the classic field of AI style robotics. The history of the origin and interchange of ideas between the different disciplines of the 'systems thinking' movement is relevant for understanding the historical background of ASR, but too complicated to be dealt with here.[14] However, *cybernetics* deserves our attention, partly because ASR can be seen as an extension of some aspects of the original research program of cybernetics, and partly because aspects of cybernetics may be viewed as a mechanist version of the functional circle of the Umwelt theory.

The idea of an art and science of control over a whole range of fields in which this notion is applicable was offered by the mathematician Norbert Wiener in 1948. Cybernetics is a theory of feedback systems, i.e., self-regulating systems such as machines and animals. The central notion is feedback, i.e., feeding back information of some change of parameters describing the state of a part of a system (e.g., some measure of output or performance) to the mechanisms responsible for effecting these changes, often with the function of regulating the system's behavior to keep it in a stable region of interaction with the environment (negative feedback).[15]

A significant example is the sensor-perception-(cognition)-motor system of our own body. When we move to catch a ball, we interpret our view of the ball's movement to predict its future trajectory. Our attempt to catch the ball involves this anticipation of its movement in determining the movement of our body. As the ball gets closer, we find it departed from the expected trajectory, and we must adjust our movement accordingly. In the cybernetic description 'we' or the subject is described as an information processing mechanism. Thus the visual system can be seen as providing inputs to a controller (our brain) which must generate control signals to cause the motor system (our muscles) to behave in some desired way (ball catching). Feedforward anticipates the relation between the system and the environment to determine a course of action; feedback monitors discrepancies which

can be used to adjust the actions. Thus, the control problem is to choose the input to the system so as to cause its output to behave in some desired way; either stay close to a reference value (the regulator problem) or to follow close upon some desired trajectory (the tracking problem). As a control signal defined by its anticipated effect may not achieve that effect, feedback is needed to compare the anticipated with the actual and to determine a compensatory change. Overcompensation gives rise to instability; undercompensation yields poor adjustment to noise and too slow performance (time delays). Thus, cybernetic principles are relatively easy to describe on the overall level, as we intuitively comprehend the principles of ball-catching in, say, baseball. Mathematically they are more difficult to analyze, and to simulate this type of behavior in full scale complexity and in real time is computationally very hard. Nobody has yet been able to design an autonomous agent that could imitate even a tiny part of the grace of human ball-catching when situated in a natural setting, e.g., a tennis court.

The cybernetic description of the information being fedback and forth between the system's components focuses on the role of the individual signs within what Uexküll called a whole functional circle. Superficially, the theoretical language of the Umwelt-theory may be translated to the language of cybernetics with no loss of meaning. However, cybernetics (as well as classical AI) is an externalist description; it does not acknowledge a subjective world of the organism experienced from within. Thus the 'information' of cybernetic feedback is not the same concept as the perceptual and operational signs of the functional circle. The latter concepts are most adequately interpreted as semiotic concepts requiring triadic relations between sign, object, and interpretants.[16] In that sense a complete meaning-preserving translation from cybernetics to Umwelt theory might not be possible; the two modes of description are partly incommensurable (but cybernetic notions can be re-interpreted and generalized in a semiotic framework). This distinction is important, because any simple device that meaningfully can be described as processing signals is information-controlled in that simple sense, even though such a cybernetic device may not have an Umwelt. It would seem absurd to ascribe a phenomenal self-world to a flywheel governor (even a fairly simple 'self').

What became of cybernetics? Today, after the introduction of computers, theoretical studies of the problems of control have become so sophisticated and their applications (to engineering, biomedicine, economics, and certainly to robotics and AI) have become so firmly rooted and self-evident that it is difficult to recapture the intellectual

excitement brought about by Wiener's ideas. Furthermore, after the cognitive revolution in psychology in the 1960s, more emphasis was put on higher level cognitive capacities which were unmanageable and intractable by purely cybernetic principles, as they seemed to presuppose the action of extensive symbolic systems for reasoning and representing information about the nature of the tasks to be solved. This initiated the whole development of CS and before that, AI. Though too crude to count as a historically valid scheme, it is not quite unfair to say that the interest in autonomous systems was represented at the beginning by cybernetics and systems science; then, in the 1950s and 1960s by the new fields of AI and robotics; and for the past thirty years by the ever changing meeting points between AI, CS, robotics, neuroscience, and recently, the Artificial Life version of theoretical biology, and ASR. This does not mean that cybernetics as such is 'dead' or that no scientific exchange or research takes place any longer under the banner of cybernetics (e.g., Heylighen et al. [eds.] 1990), but the whole field has changed with the developments of complex systems research, CS, AI, etc. Cybernetic principles are strongly integrated within the core of ASR. (As epistemology, cybernetics developed into Bateson's 'ecology of mind' or von Foerster's 'second order cybernetics', both of which are more in line with a biosemiotic study of 'the view from within'.) Let us take a look at the notion of autonomy from a more traditional *robotics* point of view, to understand what the new 'embodied cognition' and ASR movement is reacting against.

'Good Old Fashioned Robotics'

Just as AI today must be seen historically as embracing both a classical logocentric 'Good Old Fashioned AI' tradition and some more recent and theoretically broader research programs (neither of these are especially concerned with robotics), so is the term 'robotics' ambigious, and in this section we shall start to focus on a corresponding tradition of logocentric, AI-style, 'Good Old Fashioned Robotics' (GOFR).

Though there are distinct agendas within the current research programs of robotics — from pragmatic ones such as developing better attention-based motor-control systems to perform simple pre-defined tasks useful on the assembly lines in industry, to highly ambitious ones such as embodying general intelligent systems as 'servants' for human beings — the general assumption of AI-style robotics that boomed in the 1980s is that knowledge-based performance can be intelligent though mediated by a machine. In practice, AI systems may be autonomous

(as robots), or they may be intelligence-amplifiers, when used to enhance human performance in decision-making.[17] Robotic AI-systems should be capable of recognizing objects or scenes (as a real servant[18] can) and interact with worlds, that is, real worlds or computer-simulated ones (if the lab's funding is too low to take on engineers). Information for real-world interaction may be provided by television cameras and simple touch-sensors ('bumpers').

Such a complete robotic system has a learning capacity. It may learn extracting useful visual features of the information it receives and calibrate its internally represented 'visual space' by touch exploration of objects in the world. Such robotic devices[19] employ pattern recognition with stored knowledge (often represented in stable symbolic form) in order to infer (from incoming signals and the stored knowledge of their object world) the three-dimensional shapes and non-sensed properties of objects, even though the sensed data are limited and never strictly adequate. At least this was (and to some extent continues to be) the ambitious construction goal of such systems. Their pattern perception *an sich* is usually not so clever as it should be, the emphasis being on effective use of limited real-time sensed data by programs that employ stored knowledge. This generally fails in atypical situations, for inferences to the world depend on appropriate assumptions, and what is 'appropriate' is again highly context-specific and depending on the total situation while the knowledge-base of the robot is restricted to a few micro-world situations. This is an instance of the general frame problem of AI, not a minor technical nuisance but a serious obstacle to the design of any kind of system that is intended to model a complex and changing world (for details, see Janlert 1987).

Current AI research is aimed at developing programs rather than sophisticated hardware. The robot devices may serve as test beds for suggesting and testing programs if they are used in AI research at all. However, AI is not just advanced communication engineering or logic programming. Beside the goal of constructing 'intelligent' tools, AI can be (but does not necessarily have to be) presented as a claim about the nature of the mind. What John Haugeland (1985) dubs GOFAI — Good Old Fashioned Artificial Intelligence — is the strong claim that (a) our ability to deal with things intelligently is due to our capacity to think about them reasonably (including subconscious thinking); and (b) our capacity to think about things reasonably amounts to a faculty for internal 'automatic' symbol manipulation acting on a set of stable stored representations. This implies that the internal symbol manipulations must be interpreted as being about the outside world (i.e., about whatever the system deals with intelligently), and that the

internal 'reasonable' symbol manipulations must be carried out by some computational subsystem ('inner computers'). This is not just of philosophical interest, for this paradigm, when applied to the art of building robots, creates a picture of a robot as a vehicle embedding an advanced AI computer, for instance a huge expert system (where the expertise ideally should be common sense!) endowed with sensors and effectors.

In this AI-style, or, as we might call it, Good Old Fashioned Robotics, the traditional emphasis of artificial intelligence research — emphasis on explicit knowledge, rational choice, and problem solving — has proved difficult to apply to the construction of self-moving, self-orienting autonomous robots. The few systems built often show deficiencies such as brittleness, inflexibility, no real time operation, etc. The problems that have appeared in AI in this context such as the problem of non-monotonic reasoning and the frame problem (Pylyshyn 1987) are, of course, of theoretical interest and are studied in their own right, but they remain unsolved (at least within realistic time constraints), and the suggestions for solution do not appear to be particularly useful to the development of situated systems. Another characteristic of AI-style robotics is the traditional top-down design approach. None of the modules themselves generate the behavior of the total robot; one has to combine together many of the modules to get any behavior at all from the system. Improvements in the performance of the robot proceed by improving the individual functional modules. This is difficult, because of the inflexibility of the functional competence of the various parts where changes in one module will negatively affect the performance of another, so the total design has to be reconsidered in each step of design change (this problem is to some extent remedied in the new design approaches to autonomous agents). The emphases on explicitness of knowledge, rationality, and external (and top-down) design are very unrealistic from the point of view of biology and real animals' behavior and Umwelt.

Good Old Fashioned Robotics inherits *the physical symbol system hypothesis* of AI. This hypothesis,[20] which is very far from real biology, states that the processes required to produce intelligent behavior can be achieved with a collection of physical symbols and a set of mechanisms that produce a series, over time, of structures built from those symbols. The digital computer should function as a tool with which the symbol structures are formed and manipulated. Symbol structures in an AI program are used to represent general knowledge about a problem domain (such as playing chess, performing medical diagnosis, or, more relevant for autonomous robots, performing functional distinctions

between objects and creating categories of movements of the organs of a human being) and to specify knowledge about the solution to the current problem. Why should symbol systems play a necessary role in intelligent action? From the AI-type robotics point of view[21] (cf. Newell 1980), the answer seems to be that (a) rationality demands designation of potential situations; (b) symbols systems provide it; (c) only symbol systems can provide it when sufficient novelty and diversity of tasks are permitted.

Thus, the idea implicit in AI-style robotics is that perception and motor interfaces deliver sets of symbols on which the central system, or reasoning engine, operates in a domain independent way of the symbols. Their meanings are unimportant to the reasoner, but the coherence of the complete process emerges when (1) an observer of the system knows the grounding of the symbols within his or her experience,[22] or (2) the system functions so well that the complete system (the reasoning engine and the sensor-motor modules) constitutes the locus of *emergent meaning* in the sense of well-adapted functioning. Implicit in the paradigm of the symbol hypothesis is the idea that symbols and their concatenations represent entities in the world, be it individual things, properties, concepts, intentional states of other agents, perceptual qualities, etc. The central intelligence of the robot deals with symbols which must be fed into it by the perception system. It must somehow be given a (correct or approximately correct) description of the world in terms of typed, named individuals and their relationships. These assumptions are critical to the approach of Good Old Fashioned Robotics, and due to the new approaches of ASR, they are generally no longer held true.

Before we go on with this historical sketch, we should reconsider why this theoretical notion of an internal symbol system should be in contrast to the Umwelt theory? Perhaps a far-fetched question, as the two theories appear to be completely incommensurable. One might, nevertheless, interpret the situation as if AI-style robotics indeed is a hypothesis about the structure of the specific human Umwelt, which is, in some sense and to some extent, symbolic and rational. But this overlooks, first, the fact that the Umwelt theory provides a separate epistemology for the specific human Umwelt on the level of anthroposemiosis (T. von Uexküll 1986a, 1986b, 1989) that cannot be reduced to the physical symbol hypothesis, and second, the fact that the philosophical correlative to AI-style robotics is a materialist version of functionalism within philosophy of mind — the thesis that the mind is to the brain as a piece of software is to the hardware. Evidently this notion is hard to make compatible with the Umwelt theory. On the

contrary, Jakob von Uexküll's studies of the species-specific Umwelten of various animals can be seen as anticipating later 'ecological studies of perception' (the Gibson school) and notions of embodiment and situatedness in ASR (e.g., Hendriks-Jansen 1996) developed in opposition to AI-style robotics.

Biomechanical vehicles as proto-autonomous systems

People have often dreamt of building precise mechanical analogues of living beings, even if these beings were not considered to be very 'intelligent', and historical accounts of automata can tell many interesting examples.[23] In 1950 W. Grey Walter, the director of the Physiology Department at the Burdon Neurological Institute in Bristol, published 'An imitation of life'[24] describing two mechanical tortoises, Elmer and Elsie, with only two sensory organs each, and two electronic nerve cells. He called them *Machina Speculatrix*, to illustrate their 'exploratory, speculative' behavior. Historically they represent early instances of 'animats' or autonomous agents, constructed from simple cybernetic principles.

Each machine carries only two functional units, or control systems, one light-sensitive and the other touch-sensitive. With these two sense organs (an 'eye' or photocell which could scan the surroundings for light stimuli, and a simple switch sensor for touch); two miniature radio tubes; two effectors or motors (one for crawling and one for steering); and power supply via batteries, the machines could produce 'lifelike' behavior. In the absence of adequate light-stimulus Elmer (or Elsie) explores continuously (the photocell is linked with a rotating steering mechanism), and at the same time the motor drives the machine forward in a crawling motion. The two motions combine to give the machine a cycloidal gait, while the photocell 'locks' in every direction in turn. The result is that in the dark Elmer explores in a thorough manner a considerable area, remaining alert to the possibility of light and avoiding obstacles that it cannot push aside. When the photocell sees a light, the resultant signal is amplified by both tubes in the amplifier. If light is weak, only a change of illumination is transmitted as an effective signal. A stronger signal is amplified without loss of its absolute level. The effect is to halt the steering mechanism so that the machine moves toward the light source — analogous to a biological behavior known as 'positive tropism' (e.g., a moth flying into a candle). But Elmer does not go into the light source: when the brilliance exceeds a certain value the signal becomes strong enough to operate a relay in the first tube, which

has the reverse effect from the second one. The steering mechanism is then turned on again at double speed so that the machine sheers away and seeks a more gentle climate. It will circle around a single light source in a complex path of advance and withdrawal; with two light sources it will continually stroll back and forth between the two. When batteries are well charged, it is attracted to light from afar, but at the threshold the brilliance is great enough to act as a repellent so that the machine wanders off for further exploration. When batteries start to run down, the sensitivity of the amplifier is enhanced so that the attraction of light is felt from even farther away. But soon the level of sensitivity falls, the machine eventually finds itself at the entrance to its 'kennel' (a light emitting box with a certain brightness) and it will be attracted right home, for the light no longer seems so dazzling. In the kennel box it makes contact with the charger and its batteries can get recharged.

Grey Walter experimented with variations of this set-up and observed how new complex behavior could emerge in the interactions of two machines if they could sense each other (as when small lights were mounted on the shells of the tortoises). He noted that these machines, though crude, give 'an eerie impression of purposefulness, independence and spontaneity' (1950: 45).[25] Apparently these devices behave as if they have autonomous agency, and one could even ask, as we shall do below, whether they have a primitive Umwelt (though Walter to my knowledge never thought so), just like J. von Uexküll had emphasized the simplicity of the Umwelt of ticks, bugs, and other small creatures.

At that time these machines (including later modified versions by Walter) seemed to be powerful models of autonomous behavior.[26] However, during the 1950s and 1960s considerably more effort was given attempts to construct intelligent programs that could simulate higher cognitive capacities, and Walter continued his work in other directions, such as the study of brain function and stimuli association in autistic children.

In 1984 appeared a book by Valentino Braitenberg, *Vehicles: Experiments in Synthetic Psychology*[27] that became consequential not only for cybernetic-minded psychologists, but also for future work in Artificial Life and the coalescence of computational AL with hardware techniques from robotics. Braitenberg, with a background in cybernetics and neuroanatomy, describes a number of small, very simple creatures, that is, machines with sensors and motor action (mostly wheels), easily designed by simple engineering techniques, which can create often highly diverse forms of behavior. Braitenberg, who was interested in structures within animal brains that seemed to be interpretable as

'pieces of computing machinery', considered the vehicles (he did not use the term autonomous systems) as if they were animals in a natural environment. Then, one becomes tempted to use psychological language in describing their behavior, even though one knows that, according to him, there is nothing in these vehicles that their designers have not put in themselves (this is redolent of Dennett's 'intentional stance'[28]). Braitenberg makes some observations which are of general importance to the development of autonomous systems; we shall briefly consider the most important.

The first point is about the kind of physics in which the vehicle 'lives'. A vehicle must be able to move, but even planets move, so what is special to the movement of an autonomous system? Braitenberg describes the simplest species — Vehicle 1 — as equipped with just one sensor in the front and one motor in the back of the device with a very simple sensor-motor connection (Figure 1).

The more there is of the quality (e.g., temperature) to which the sensor is tuned, the faster the motor goes. The vehicle moves in the direction it happens to be pointing; it will slow down in cold areas and speed up where it is warm. But it lives on the ground (or in water), that is, in a world in which Newton's law of inertia does not make direct sense; rather, it is a world of friction, an Aristotelian world in this sense. Friction slows down the body, and if the vehicle enters a cold region where the force exerted by its motor, being proportionate to temperature, becomes smaller than the frictional force, it eventually comes to rest. Now Braitenberg asks us to imagine a vehicle of this kind swimming around in a pond: 'It is restless, you would say, and does not like warm water. But it is quite stupid, since it is not able to turn back to the nice cold spot it overshoots in its restlessness. Anyway, you would say, it is ALIVE, since you have never seen a particle of dead matter move around quite like that' (Braitenberg 1984: 5).

By a kind of incremental methodology Braitenberg increases the complexity of a series of vehicles, Vehicle 2 being just a kind of duplication of the first one, with two motors and two sensors in the respective corners of the trunk, and coming in two distinct varieties depending on whether the right sensor is connected to the right motor and vice versa or they are cross-connected (see Figure 2). If there is no crossing, the

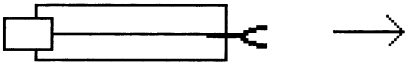

Figure 1. *Vehicle 1. The leftmost little box is a motor organ; the body with the sensor-motor connection is in the middle; the Y fork is a sensor; and → indicates the direction of movement*

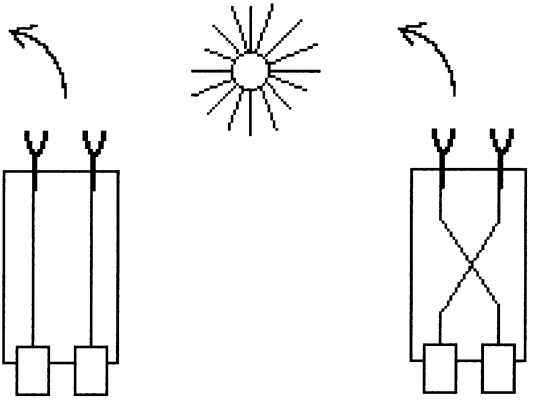

Figure 2. *Vehicle 2*

motor on the side of the body, which gives the highest exposure to the sensor of the stuff that excites the sensor, will tend to move faster, so that as a result, the vehicle will turn away from the source (it will 'fear' it, as Braitenberg says). In the vehicle with crossing the resulting movement will turn the vehicle toward the source (indicated by the sun icon) and will eventually hit it.

This is, of course, just the beginning; and Braitenberg developed a whole series of agents with a wide spectrum of capacities which he interpreted from his intentional stance as showing 'fear', 'aggression', 'love', etc. We shall not dwell on the details, because the importance to ASR is clear. Even though cybernetic principles could not serve as a basis for 'cracking the cognition problem' (e.g., constructing systems that would show general intelligence, general problem-solving capacity, planning, etc.), the construction of biomechanical vehicles revealed that simple behaviors of simple machines may, in a varied environment and interacting with other machines, produce something which looks like organisms governed by quasi-intelligent control-structures. Like true living animals, they seem to constitute simple functional circles of semiotic processes of sign-interpretation and sign-action. Why should they not have Umwelten?

The various attempts to build biomechanical autonomous[29] vehicles differ in one important respect from the earlier clock-work automata (e.g., the 'drawing' automaton of the Jaquet-Droz family, cf. Chapuis and Droz 1958) as well as the AI type of robots that followed the cybernetic period: *They did not rely on a central 'program'*. Robots and nineteenth-century automata did that: The program was responsible

for the model's dynamic behavior. Whether it was a rotating drum with pegs triggering levers in sequence, a set of motor-driven cams, or some other mechanism, the movement that the automaton realized was 'called for' by a central control machinery. As emphasized by the new movement of complex dynamical systems research,[30] therein lay the source of failure of these models and the limited perspective of a whole program of modeling human and animal systems that followed right up to (and included) much of the work in AI. The most promising approaches to model complex phenomena like life or intelligence became those which dispensed with the notion of a centralized global controller, and focused instead on the mechanisms for the distributed control of behavior, situated activity, and the types of emergent dynamics that form out of local interacting agents. Grey Walter and Valentino Braitenberg did a pioneering job to model autonomous systems without presupposing elaborate and explicitly encoded control structures. In a biosemiotic perspective it is a historical irony, or at least interesting, that both such a relatively 'mechanical' paradigm as cybernetics, and 'the vitalist' Jakob von Uexküll's idea of the Umwelt, with its emphatic notion of the unity of the organism and its sensed environment, must be seen as precursors to the recent concepts of situatedness and embodied cognition developed in the context of ASR. Let's take a look at that research.

Autonomous agents

During the 1990s new ideas, concepts, and methods have been used to re-vitalize the best of the old cybernetic approach to simple robot design (also revived by Braitenberg on a more simple scale); a set of projects that became known as design of autonomous systems, agents, or animats. Already from the middle of the 1980s, Rodney A. Brooks and his group at the Massachusetts Institute of Technology's Artificial Intelligence Laboratory developed a practical critique of the 'Deliberative Thinking Paradigm',[31] that is, the thesis that intelligent tasks can be (and always have to be) implemented by a reasoning process operating on a symbolic internal model; which was described as Good Old Fashioned AI and Robotics above. This critique, together with a new set of modeling techniques and construction principles, gradually gained in influence[32] and became known as the 'reactive systems movement', 'agents research', etc. Closely parallel to this movement and deeply inspired by it, new notions of cognitive processes as being enacted in embodied situated systems were

developed.[33] Interestingly, one of the sources from which this new paradigm of robot construction and embodied cognition was inspired, was Jakob von Uexküll's Umwelt theory as it was used to emphasize, first, the tight dynamic connection between the animal's body plan and its experienced world, and second, that the world perceived by the animal was totally different from the world perceived by the zoologist, indicating the need for an increased awareness of the fact that a robot would live in a 'perceived' world that differed much from what the robot builder immediately might be able to see.[34]

The principles of agents design

Brooks' group and other researchers found it unrealistic to hope that more action-oriented tasks could be successfully implemented by a 'deliberative' machine in real time, so they started to develop new ideas about how autonomous agents should be organized; a project that, according to Maes (1990) led to radically different architectures. These architectures (as discussed in, for example, Maes 1990; Brooks 1986a [same argument as in Brooks 1991b]; Brooks 1992; Meyer and Guillot 1991; Wilson 1991; Brooks and Maes 1994; Clark 1997; Ziemke and Sharkey 1998) are generally characterized by

- emergent functionality
- task-level decomposition
- more direct coupling of perception to action
- distributedness and decentralization
- dynamic interaction with the environment
- physical grounding (situatedness and embodiment)
- intrinsic mechanisms to cope with resource limitations and incomplete knowledge

The *functionality* of an agent is considered as an *emergent* property of the intensive interaction of the system with its dynamic environment.[35] The specification of the behavior of the agent alone does not explain the functionality that is displayed when the agent is operating. Instead, the functionality is to a large extent grounded in the properties of the environment. What seems to be a complex behavior does not necessarily have to be coded in the agent; it can be an outcome of a few simple behavioral rules and the interaction with the environment. The environment is not taken into account predictively or dynamically, but its characteristics are exploited to serve the functioning of the

system. Thus, one cannot simply 'tell' these agents how to achieve a goal. One has to find an interaction loop involving the system and the environment which will converge (given that the environment has the expected properties) towards the desired goal (this sounds perhaps easy, but, in fact, often proves to be hard to 'find' such a loop).

The second feature is *task-level decomposition*. This does not mean the same as task decomposition in classical AI. An agent is viewed as a *collection of modules* each of which has its own specific domain of interaction, or competence. The modules operate quasi-autonomously and are solely responsible for the sensing, modeling, computation or reasoning, and motor control which is necessary to achieve their specific competence. The agent design approach does not abstain from using representational notions or AI-reasoning techniques, but the conceptual framework within which these notions are seen has changed, because there is no central reasoning module that plans or governs the overall behavior, nor any global planning activity within one hierarchical goal structure. To avoid costly and unnecessary duplications of the modules, they may make usage of 'virtual sensors'. Communication among the modules is reduced to a minimum and happens not by using high-level languages, but on an information-low level. The general behavior of the agent is not a linear composition of the behaviors of its modules, but may emerge by the interactions of the behaviors generated by the individual modules.

The *direct coupling of perception to action* is facilitated by the use of reasoning methods which operate on representations which are close to the information of the sensors (i.e., 'analogical' representations[36]). If a problem such as categorization of objects can be solved by processes dealing with sensation or perception rather than symbolic cognition, this is preferred. Perception may be made less general though more realistic; there is no need for the perception system to deliver a description of the world as in the AI-style robots. The special 'subsumption architecture'[37] enables the designers to connect perception more tightly to action, embedding robots concretely in the world, to use another popular phrase of this approach. Again, we may ask: Why not see this as an attempt to develop a more specific theory of the internal workings of an Umwelt? We shall soon return to this question.

The agents approach or 'nouvelle AI' is based on *the physical grounding hypothesis*. It states that to build a system that is intelligent it is necessary to have its representations grounded in the physical world.[38] What exactly this means is seldom made fully explicit, but some hints of the idea can be given. A physically grounded system is one that is connected to the world by sensors and actuators/effectors, in a functional

circle as it were. Thus it is not adequate to study, for example, problems of perception merely by simulation techniques; typed input and output are no longer of interest because they are not physically grounded. Intrinsic to the idea is also that systems should be built in a bottom-up manner. High level abstractions have to be made concrete. The constructed system has to express all its goals and desires as physical action (as opposed to stored non-dynamic representations in the memory); and the system should extract all its information from physical sensors, i.e., the initial input should not be delivered to the systems as symbolic information, but rather as physical action. The designer of such a system is forced to make much more design components explicit. Every shortcut has a direct impact upon system competence; there is no slack in the input/output representation. Often the very notion of representation as something explicit and stable is criticized.[39] (This has even led some researchers to an 'antirepresentationalist view of cognition', which is, however, an inadequate way of expressing the fact that GOFAI had a restricted and simplistic view of such categories as 'representation' and 'symbol'; one should rather reconstruct various kinds of representation in various kinds of systems as a continuum of cases within a general semiotic and triadic model of representation, as suggested by Katz and Queiroz 1999.) A slightly different way to state the idea of physical grounding is by the notions of situatedness and embodiment (Brooks 1991a,b; cf. Hendriks-Jansen 1996). *Situatedness* implies that the robots are situated in a world; they do not deal with abstract descriptions, but with the here-and-now of the environment that directly influences the behavior of the system. *Embodiment* implies that the robots have bodies and experience the world directly and that the actions have an immediate feedback upon the robot's own sensations.[40] Computer-simulated robots may be 'situated' in a virtual environment, but they are certainly not embodied.

Life and intelligence: The perspectives of agents research

One might ask, of course, if these requirements are sufficient to secure that a system so constructed with such and such behavioral capacities is *intelligent* (in the rationalistic Newell-sense of being 'general intelligent')? Probably not! But here, one must notice a crucial difference between Good Old Fashioned Robotics and the new approach with respect to the apprehension of the concept of intelligence. According to the classical approach, intelligent behavior presupposes the capacity for rational manipulation of elaborate internal symbolic

structures — a 'language of thought' of some kind — representing the state of affairs in the real world. Though a language of thought need not be used for linguistic communication, it is thought that very few (if any) animal species can have a representational capacity of the same order of magnitude and complexity as that of the glossophile *Homo sapiens*. On the other hand, researchers in situated activity, agents, and Artificial Life agree that many animals are 'intelligent' to some extent. The factual evolution of intelligent animals is considered to be an instructive standard to understand the basic requirements of intelligent behavior. Computationally, the most difficult things to achieve by evolution seem to be the ability to move around in a dynamic environment, and to process sensory information in adaptive ways to ensure survival and reproduction. This part of intelligence is where evolution has concentrated its time; the physically grounded parts of animal systems.[41] From the perspective of the Umwelt theory, we can see these parts as closely related to the emergence of complex Umwelten. So the primary Umwelt evolution is computationally costly so to speak: it takes many evolutionary time steps. This is also the case for simpler life forms such as unicellular eukaryotic cells (*Protozoa*) which lack a nervous system and a genuine Umwelt, but do (according to T. von Uexküll 1986a) have a simpler 'autokinetic or self-moving circle' by which they enter into semiotic interactions with their exterior milieu.

The evolutionary perspective of 'nouvelle AI' seems to be promising. A growing group of AI-specialists recognize the limitations of the purely logical approach to constructing thinking machines and are attracted by biologically-inspired design principles which may form a basis for the architecture of hardware and software in computers of the future.[42] Biologically inspired AL techniques serve as inspiration for finding more natural ways of viewing the design problems of robotics. Organisms were not released into nature only after they were constructed as functionally perfect designs; evolution operates like a tinker who fixes a broken machine using the materials at hand. Not every design is a good design, many are tried out in evolution but quite few basic types survive. The robot builders may learn something by studying the evolutionary game. Instead of constructing expensive, complicated machines designed for a limited number of well-defined tasks, one might instead build, to follow Brooks' advice, a whole flock of cheap, simple, and perhaps rather unpredictable machines and allow them to evolve gradually.

For cognitive science the new ASR, or 'Behavior Based AI' movement, may lead to a considerable change of perspective. Perhaps one might not be able to 'crack the cognition problem' or create an understandable

scientific theory of thought until we understand what it means to say that something is alive. Life came before real intelligence, and autonomous systems and A-Life should come before real AI. The problem with AI research may be that one has sprung directly into the most complex example of intelligence — human intelligence. It is tempting to suspect that we have been cheated by the fact that computers can do some things which people find difficult. ASR, AL, and AI are in a process of being transformed into a continuum of projects which attempt to model adaptive, learning, and cognitive abilities in all the varying degrees of complexity we know from biology and psychology. ASR can be viewed as a science which concerns itself with the minimal level for thinking and the lower limit for sign-manipulation and 'computation': how simple must a physical system be before we cannot anymore call it computational (cf. Emmeche 1994b) and alive? Or, phrased in terms of the Umwelt theory: What is the minimal (artificial or natural) system realizing its own Umwelt?

This 'AI must be AS thesis'[43] (or 'intelligence demands autonomy') can be formulated as follows: 'The dumbest smart thing you can do is to stay alive' (Belew 1991). Animals do it, and humans do it, too. Intelligence does not concern either/or, but different ways of managing the requirements of self-maintenance and adaptation. Organisms' apparently highly evolved, coherent behavior can often be explained from the quite simple reciprocal interactions with a rich and varied environment. Much of the complexity seems to lie in the milieu. Think of an ant.[44] It crawls around on the forest floor, carefully avoiding large barriers but must take minor detours in order to get space for dragging home a pine needle to the ant nest. The ant pauses from his work and exchanges information with a fellow ant. It usually has an especially complex route, so it seems. But the ant as a behavioral system, as well as its Umwelt, is quite simple — the complexity is to a great degree a reflection of the environment in which it finds itself. The point here is that if we have visions of constructing serviceable, sociable robots or such things, we must first discover the minimal procedures which enable an animal to cope minimally with its nearest surroundings. This does not sound like much, but it is! An ant can never imagine what it meets on its path. Openness, adaptability, and flexibility become more important than having a ready response to every conceivable situation, regardless of whether the response can be coded as a frame, a scheme, a script, or as one of the other AI-techniques of representing knowledge. Thus it appears that Umwelt, autonomy, 'intelligent' action, and embodied knowledge are closely coupled. But is this the whole story? Have some rather hard problems been left out?

Is anybody home? Can Umwelten be artificial?

When asked if it is possible for these artificial (human constructed) robot-like animats or autonomous systems to have an Umwelt, people seem to have two quite different intuitions. One of these can be expressed in the answer 'Yes, why not? If such a simple living creature as the tick does have an Umwelt, even a very simple one, why not the robot?' — a quite reasonable answer it seems. Another, opposing answer would be: 'No, of course not! How foolish! It's just a piece of electronics. No matter how complicated the circuits of its artificial neural network are (or whatever intermediates sensor and motor modules), how could you even think it should feel anything at all?'. Personally I would immediately and intuitively tend to the no-Umwelt-in-a-robot answer, though I can certainly follow some arguments for the yes-there-is answer. But as intuitions simply divide people and may be misleading, let us look at some arguments.

The robot-does-have-an-Umwelt answer can be stated like this: Premises: 1) All it takes to constitute an Umwelt in the sense of a phenomenal experienced species-specific (or 'device-specific') world is a certain circular information-based relation between sensor devices and motor devices as described by the notion of a functional circle. 2) Clearly, even simple artificial animats (like Grey Walter's Elmer) instantiate such a circle, just like simple animals do. 3) Conclusion: Artificial autonomous systems such as robots do have an Umwelt (from which it per definitionem follows that for the robot there must be something it is felt like or experienced like to be — just as there is for me, a dog, or a tick).

The no-Umwelt-in-a-robot answer, acknowledging that the robot indeed instantiates a functional circle in the sense of a causal feedback loop, does not hold this circle to be a true instance of a functional circle in the semiotic sense of forming, by sign action, the backbone of an experienced Umwelt. Why not? Because from this perspective, what gives the Umwelt its phenomenal character is not the functional-cybernetic aspect of signal-processing within the system (and at the system-environment interface), but the fact that the living organism is beforehand constituted as an active subject with some agency. Thus, only genuine living beings (organisms and especially animals) can be said to live experientially in an Umwelt.

Here the counter argument would be that the no-Umwelt answer seems to presuppose what that argument should show by putting the criteria for the existence of an Umwelt inside the agent as a kind of hidden (occult!) capacity, incidentally only found in some kinds of devices,

viz. the organic non-artificial ones, instead of allowing for an objectively accessible behavioral criterion for the existence of an Umwelt (e.g., the existence of what must be described as information processing within some given type of functional architecture). Thus the no-Umwelt answer is really not an argument, it is a restatement of an a priori intuition.

In a sense this counter is fair, but the presumption that only objectively accessible behavioral criteria count as criteria for anything we can identify and study scientifically is an externalist presumption which does not necessarily hold true for a rational understanding of a certain range of phenomena (intentional phenomena, qualia, consciousness, etc.), at least from the point of view of some traditions of scientific inquiry which are not exclusively externalist (e.g., semiotics, phenomenology, hermeneutics). Furthermore, the Umwelt-in-a-robot answer presumes that to assign an informational-cybernetic description of the device's dynamics is trivially the same as identifying and explaining the existence of a robot's Umwelt as an intrinsic phenomenon — which is hardly convincing, as it would imply that even simpler cybernetic devices (as the fly-wheel governor) should realize an Umwelt.

Before resolving this issue, we have to look closer at (a) the subject-character of the Umwelt according to the very Umwelt concept; (b) the semiotic aspect of the Umwelt and its dependence on qualitative aspects of sign action and sign interpretation (especially the notion of qualisign); (c) the possible realizability of sign action in non-organic media; and (d) the general epistemic non-accessibility of (at least certain qualitative aspects of) the Umwelt by those other than its owner. Finally, (e) we will discuss various sorts of 'situatedness' in ASR and the artificiality of robot situatedness.

(a) If what it means to have an Umwelt is to be an active subject with some agency, we should keep in mind that the way such a thing as an Umwelt exists (according to the definition of the Umwelt given above) is ontologically different from the way the physical environment (as studied by ecology) exists, or the way the neural system as a complex dynamic biophysical network (as studied by neurobiology) exists, or the way the observable behavior of the animal (as studied by ethology) exists. To say that it is subjective means exactly that it exists in the mode of an active experiencing subject, which is not something that can be seen or described from a purely external point of view (cf. T. von Uexküll 1982a; Nagel 1986; Searle 1992).

(b) That this subjective aspect of animal sensation, perception, and even cognition is accessible for a semiotic description is because of its very character of being based upon triadic sign relations, which are not just adaptive (and thus biologically meaningful from the Darwinian

functionalist perspective of survival), but truly significant (in the inner experiential sense) for the animal in question. A sign may have all sorts of relations to other signs and all sorts of effects in its process of being interpreted by the system of interpretance (in this case the organism), but according to Peirce's sign notion even more developed signs — such as symbols (and the arguments of humans) — include within them simpler ('degenerate') signs, in which the aspects of secondness and firstness are more prominent. That is, the internal signs mediating an Umwelt's *Merkwelt* and *Wirkwelt* (i.e., mediating the perceptor and motor/operator organs) do have a qualitative aspect to them, an aspect that is often overlooked both by semioticians and biologists. A single sign may be a token of some general type (e.g., a perceived pattern may be recognized by the organism as being of a certain dangerous kind, say, a predator), but it has always also an aspect of being a tone, i.e., being qualitatively felt in some way (e.g., unpleasant). The tone/token/type is a genuine triad, where the firstness property of the tone is always partly hidden, so to speak, within the 'objective' or more external property of that sign's belonging to a type. This corresponds to the first trichotomy of signs in Peirce (that trichotomy according to the character of the representamen itself), where every legisign is always realized by a particular sinsign, and every concrete sinsign involves a qualisign.[45] One should remember that the different kinds of signs in Peirce's classification are not isolated distinct entities but have specific internal relations, such as an inclusive relation of the higher categories of signs including or presupposing the lower ones.[46] What is a qualisign? Only phenomenologically can we approach a clear idea of the qualisign; it is of an experiential character, it is, as Peirce says, 'any quality in so far as it is a sign', 'e.g., a feeling of ''red'' '.[47] Thus, the Umwelt, as a semiotic phenomenon, includes qualisigns with very sensuous 'tonal' qualities. (That semiosis is generally a phenomenon of thirdness does not mean that the qualitative firstness of signs is absent.)

(c) Could qualisigns be realized in artefacts, devices designed by humans? I do not believe that to be the case, but I think it depends. From a Peircean point of view, this might be the case (at least potentially), but it depends upon 1) the semiosic capacities of the constituent materials to realize habit formation and continuous living feeling and 2) the organization of the very device. Why are the materials important? Isn't that carbon-bio-chauvinism? I have argued elsewhere that in biological cells, the sign-aspects of their internal actions are not medium-independent, that is, the process-structure of the 'information' in the cell can only be realized by the cell's highly specific biomolecular materials (Emmeche 1992). If such a material device as a robot could have that

special organic flexibility of an animal that allows it to instantiate anything like the law of mind, that is, the tendency to let signs influence or affiliate with other signs in a self-organizing manner, it is difficult to see why such devices should not be able to realize genuine signs (including qualisigns) — and thus not simply be systematically interpretable as doing so by an external human observer (which is the usual case for robots: they are easily interpreted as being intentional agents, but taking this 'intentional stance' by their constructors does, of course, not tell anything about the eventual existence of their own 'feelings'). If that artificially constructed system realizes qualisign action and living feeling, it would have mind in Peirce's sense. But would it have an Umwelt? Is there anybody home in there, experiencing something? Remember the extremely broad scope of semiosis and mind in Peirce's sense. If the very universe anyway is perfused with signs, according to Peircean pansemiotics (as in Merrell 1996), this state of affairs may not help us to decide whether a robot is experiencing anything; whether it has an Umwelt. It might have, we could imagine, if the qualisigns and all its higher forms of semiosis became organized in such a way as to make possible the emergence of that sort of unity and experiential coherence that characterizes 'an Umwelt-as-we-know-it' (our own).

(d) But how should we know? An Umwelt can only be directly known from within. Does the general epistemic non-accessibility of any Umwelt by those other than its owner imply that when we face an actual or future robot, we are exactly in the same situation as facing a tick, a snake, humming bird, or a dog? They do have an Umwelt, but how this Umwelt really is and how it is felt like is impossible to say.[48] These two situations are parallel but not the same. In the case of any living animal with a central nervous system (including a tick!), we can be rather sure that they do have an Umwelt somehow. The machine view of life and of other people has been transcended. Philosophically, the only solution to the so-called 'problem of other minds' (how can we be sure that other persons do have a mind?) is either to say pragmatically that this is no problem at all, or to say, because I know I have, and they are similar to me, by analogy they must have too, and it is the best explanation of their behavior that it is indeed connected to and partly explainable by their minds. The analogy solution is also what we do (often less convincingly) with animals. We know that 'there is someone home' in the dog's organism, though the actual content of its mental states is harder to infer. Even harder for the snake, and so on. But this everyday analogical inference is, in fact, backed up by biology. The bird's brain as an organ is indeed not only analogous in the biological sense (same overall function) to our brain, it is homologous to it (it has the same evolutionary origin).[49]

We all descend from very simple creatures which did have basically the same kind of organ called a nervous system (NS), including a brain. So in the animal case, the 'problem of other animal's Umwelt' is answered by a combination of an externalist explanation of the NS-homology (where NS plays the role of a necessary condition for an Umwelt), and an internalist knowledge of one's own experiential Umwelt, plus the analogy inference just mentioned (supported by the general Umwelt theory). But in the robot case, the 'problem of a robot's Umwelt' is different. It cannot be posed as a 'problem of another machine's Umwelt', because we are not machines (cf. Kampis 1991), that is, we cannot use the analogy inference; nor can we appeal to evolutionary homology. So even if a robot behaves like a fully autonomous system, the inference that the robot then does have an Umwelt is not warranted by these arguments.

(e) Does that mean we should distinguish between true 'situatedness' for animals and artificial robot situatedness in the context of ASR and AL? Those who have enthusiastically drawn parallels between 'situated cognition' in robots and in humans (cf. Hendriks-Jansen 1996; Clark 1997) may still miss some crucial qualities of true biological cognition. One interesting possibility comes to mind: That only systems (whether robots or organisms) which do have an experiential Umwelt could realize such complex behaviors as we see among higher vertebrates and in humans.[50] If the Umwelt is a higher level emergent phenomenon, emerging from the embodied sign processes in the nervous system of a situated agent, a necessary condition for the system's graceful performance might well be the constraints of the Umwelt upon the particular patterns of movement, a kind of 'downward causation'.[51] In artificial neural networks, just as in computational Cellular Automata, all the causal efficacy of the dynamics of the system seems to be located at the low-level rules of input-output behavior of the individual components, thus what appears to be emergent for an observer may, in fact, just be emergent in the eyes of that beholder (Cariani 1992). In contrast, for the intrinsic emergent perception and cognition in animals and humans, it might be the case that 'mind over matter' is not just a metaphysical speculation, but, when transformed within a general Umwelt theory of dynamic levels of interaction and semiotic levels of interpretation, becomes a true principle of self-organization and downward causation in complex systems (cf. Lemke 2000). This is certainly the position of the biosemiotic variant of qualitative organicism: The emergent phenomena experienced as components of the system's Umwelt play a direct role in the behavioral dynamics of that living system: they constitute a complex system of interpretance (cf. Sebeok's internal modeling system) that

continually shapes the individual movements of the animal. Here, we are reminded of the old Aristotelian idea that the animal soul and genuine animate motion is one and the same phenomenon (cf. Sheets-Johnston 1998, 2000). Also implied here is a strong notion of the qualitative complexity of certain systems. According to this notion a system is *qualitatively complex* if (i) it is self-organizing, (ii) it has an Umwelt with experiential qualia, and (iii) a condition for (i) is having (ii) — which means, that in order to have the capacity for a process of self-organization, and on the behavioral level a high-order graceful behavior, the system has to have some aspect of qualitative experience. That is, the Umwelt has somehow the causal powers of organizing (by downward causation) the total 'self' of the system, to make it cohere, to give it its form of movement. The very notion of a qualitative experiential self may be interpreted as an emergent property of the parts of the system and their dynamics, including the organism-environment interactions and the genealogical trajectory of the system over time: The self can be interpreted as an interactional, situated, historical, and emergent self, the very agency aspect of a living system. This notion of complexity cannot be reduced to any linear increase of a quantitative measure.

In the realm of agents moving through a changing and challenging environment we can formulate the difference between artificially situated robots and truly situated animals with an Umwelt as the difference between, on the one hand, to be able to look where one is going, look out, watch one's steps — all of which an autonomous system has to 'learn' if it should act like an agent — and on the other hand, to be able to behold something in one's experiential world, see and feel it with one's 'inner eye'. The big claim (not yet proven and maybe in principle unverifiable) of qualitative organicism is that having an experiential Umwelt is a precondition for really being able to achieve full-scale autonomy with all its gracefulness of movement that only higher animals have yet achieved. If the artificial systems are only partly 'situated' as they do not in the full sense of the word experience an Umwelt, there is indeed some hope — by a closer approachment of theoretical biology, semiotics, autonomous systems research, and cognitive science — for gaining a deeper understanding of truly situated autonomous systems as being a kind of complex self-organizing semiotic agents with emergent qualitative properties.[52]

Notes

1. As for the Umwelt part of this distinction, one could further demarcate between the Umwelt in a more narrow sense as the species' significant surround and the

Innenwelt as an individual organism's actual version of that surround (cf. Anderson et al. 1984: 13); but this distinction is not necessary in the present context. Innenwelt does not figure in J. von Uexküll (1940), or in T. von Uexküll's glossary to this text.

2. Needham suffices to serve as an example here: 'Today we are perfectly clear … that the organisation of living systems is the problem, not the axiomatic starting-point, of biological research. Organising relations exist, but they are not immune from scientific grasp and understanding. On the other hand, their laws are not likely to be reducible to the laws governing the behaviour of molecules at lower levels of complexity' (from the l937 essay 'lntegrative levels: a revaluation of the idea of progress' in Needham 1943).
3. On British emergentism, see Beckermann et al. (eds.) (1992); on the role of the notion of the organism as a special emergent level of integration, see Salthe (1988).
4. Kant's philosophy of biology in *Kritik der Urteilskraft* was a significant source for Jakob von Uexküll.
5. Conceiving a mathematical basis for mechanics, Galileo (1564–1642) in *Saggiatore* (1623) elaborated a distinction recommended by Democritus (c 460–371 BC) between those qualities inherent in or produced by inorganic bodies (shape, size, location in space and time, and motion) and all other qualities which are in the observer, not in Nature (heat, sound, taste, etc.). Robert Boyle (1627–1691) later called this radical demarcation primary and secondary qualities, a distinction Locke (1632–1704) systematized.
6. Though one should not put individual thinkers in categories that do not adequately express their views; just to give a very rough indication of possible representatives of the two positions, *mainstream organicism* is often expressed within such heterogeneous currents as 'classical' neo-Darwinism (E. Mayr, etc.), 'dynamical' Darwinism focusing on self-organizing systems and selection (S. Kauffman, D. Depew, B. Weber), artificial life and autonomous agent approaches (C. Langton, R. Brooks, etc.), the developmental systems approach (S. Oyama, P. Griffiths, E. Neumann-Held), the morphodynamic field approach (B. Goodwin), and hierarchical conceptions of evolution (S. Gould, N. Eldredge, S. Salthe, etc.). *Qualitative organicism* is represented by biosemiotics (J. Hoffmeyer, T. Sebeok, J. and T. von Uexküll, K. Kull, etc.), 'the animate approach' (M. Sheets-Johnston), the notion of a biological science of qualities (B. Goodwin), and to some extent also by studies of animal communication from the point of view of an 'ecology of mind' (G. Bateson), and even 'activity theory' deriving from the Soviet cultural history school (Luria, Vygotsky, etc.). I find the Second Order Cybernetics (H. Von Foerster, G. Pask, etc.) and the 'internalism' (see Van de Vijver et al. 1998) of thinkers like K. Matsuno, S. Salthe, and others more difficult to place, but it is probably related to qualitative organicism as well.
7. It is often assumed that to the extent that these subjective aspects of animal life, say pain, can be seen as subserving survival of the organism, they do have a functional adaptive explanation within a neo-Darwinian frame of evolution by natural selection. This is not so. Selection cannot 'see' the pain of an animal. The animal could, from the point of view of the selective story, just as well be an insentient zombie that had preserved the same functional input-output relation between, say, detection of inflicting actions upon the organism and its adaptive withdrawal from such actions which caused eventual detriments. The neo-Darwinian explanation scheme is a completely externalist approach and cannot account for the internal experiential world of the animal. There is as such no reason why highly organized information processing Darwinian devices should feel anything at all.

8. The accusation for carbon-chauvinism could not be directed against such disciplines as ecology or ethology; it only became possible as a result of the 'molecular revolution' in biology after the Watson and Crick discovery.
9. For some critical voices, see Pattee (1989); Kampis (1991); Cariani (1992); Emmeche (1994b); Moreno et al. (l997).
10. Although similar ideas may be introduced, e.g., Rasmussen (1992) who uses John Wheeler's 'meaning circuit' to postulate that an artificial organism must perceive a reality of some kind. Interestingly, Sebeok (1991) relates Wheeler's as well as J. von Uexküll's ideas to 'the doctrine of signs'.
11. In moral philosophy, the term is used in this sense (e.g., Mele 1995), a source of potential confusion.
12. An example of a biological use is the designation 'the autonomic nervous system', that is, the system of motor (efferent) nerve fibers supplying smooth and cardiac muscles and glands (comprising the sympathetic and parasympathetic nervous system), which is not 'controlled by will' (of the autonomous person) but is self-governing.
13. Here, the concept of organization is considered as the relations that define a system as a unity and determine the dynamics of interactions and transformations that the system may undergo; the organization of a living system is considered as autopoietic.
14. For a history of cybernetics, see Mayr (1970), of systems thinking see Lilienfeld (1978).
15. An example is a machine in which we distinguish four parts: a flywheel W, a governor G, a fuel supply device F, and a cylinder C; this is, as one can see, framed over the James Watt 'governor' invented to regulate the velocity of rotation in steam-engines, where the output (rotation velocity) regulated the input (the steam). The machine is connected to the outside world by the energy input and the 'load' which is considered as a variable and weighing upon the W. The central point is that the machine is *circular* in the sense that W drives G which alters F which feeds C which, in turn, drives W. How does it work? The more F, the higher velocity of C. The higher C, the higher speed of W. And, as the feedback is negative, the higher speed of W, the lower supply of F. (If the feedback were positive — if, for instance, higher speed of W caused higher supply of F — the machine would go into a *runaway*, operating exponentially faster until some parts might break). The example is due to G. Bateson, who was much inspired by cybernetic principles in his attempt to develop a 'mental ecology'. He notes (in Bateson 1979 [1980: 117]) that in the 1930s when he began to study these systems, many self-corrective systems were already known as individual cases, but the cybernetic principles remained unknown. He lists among the individual cases Lamarck's transformation (1809), Watt's governor (late eighteenth century), Alfred Russel Wallace's perception of natural selection (1856), Clark Maxwell's analysis of the steam engine with a governor (1868), Claude Bernard's *milieu interne*, Hegelian and Marxian analyses of social processes, and Walter Cannon's *Wisdom of the Body* (1932). — One might add Felix Lincke's lecture from 1879, *The Mechanical Relay*, which was probably the first attempt to outline a unifying theory of feedback control applicable to machines as well as to animals (cf. Mayr 1970).
16. A detailed Peircean re-interpretation of J. von Uexküll's functional circle is not my principal aim here (though see below on qualisigns in the Umwelt), but Thure von Uexküll has gone some way in that direction (e.g., Uexküll 1982b), see also Hoffmeyer (1996) and Salthe (this issue). Note that the Peircean notion of representation (cf. Nöth 1997) is both very complex, general, and dynamic, and cannot be equated with the simplistic AI idea of representation as a direct mapping between internal symbols and external objects (cf. Figure 2 in Ziemke and Sharkey, this issue).

17. The notion of 'intelligence-amplifier' is, of course, vague, because such aids as slide rules, pocket calculators, or even paper and pencil may be regarded as 'intelligence amplifiers' for humans though they are not intelligent (Gregory 1981). The strong claim of robotics and ASR that intelligence may be realized artificially can be formulated in this way: To the extent that these devices really are autonomous, their intelligence is 'intrinsic' to them, it is not derived from human intelligence or merely ascribed to the system. This, of course, begs the question of what it means to be 'really autonomous' (Is it simply the capacity to function for some time without human intervention? Is it the ability to move around and orient 'oneself' in an environment and solve simple problems? Is it being an autopoietic system? Or is it, for instance, the capacity to go on and live a life of one's own, reproduce, and thus contribute to the maintenance of an evolving population?).
18. A commentator on AI and robotics once remarked that the major goal of this species of research seems to be synthesizing 'the lost mother' as an eternal and non-demanding servant to care for you and do all the tiresome practical work that your own mother used to do for you when you were a child. This, in fact, is another definition of a full-fledged autonomous system: An artificial mother that keeps you going!
19. Traditional AI-type robotics is surveyed in, e.g., Gevarter (1985); see also Pylyshyn 1987.
20. See Newell (1980) who formulates this hypothesis thus: 'The necessary and sufficient condition for a physical system to exhibit general intelligent action is that it be a physical symbol system' (170); where the physical symbol system is a 'universal machine' which is physically realizable; 'any reasonable symbol system is universal (relative to physical limitations)' (169). Newell defines 'universal' with reference to Church's thesis (also called the Church-Turing thesis). He clearly states that the advances in AI (such as reasoning and problem solving) 'far outstrip what has been accomplished by other attempts to build intelligent machines, such as the work in building robots driven directly by circuits …' (171). Brooks (1990), in his critique of the thesis as paradigmatic to AI-type robotics is more sloppy: 'The symbol system hypothesis states that intelligence operates on a system of symbols'.
21. However, from the perspectives of new robotics (Autonomous Systems Research, see below) one may question the extent to which theoreticians like Simon, Newell, Fodor, and Pylyshyn actually ever cared much about the real hard theoretical and practical issues involved in robot building.
22. This is the only alternative Brooks (1990) sees in his critique of the symbol system hypothesis. Notice that this is the important 'symbol grounding problem' (Harnad 1990), see also the series of papers by Stevan Harnad referred to in Hayes et al. (1992).
23. A famous European example of such an early man-designed 'autonomous' system is Jacque de Vaucanson's mechanical duck from 1735; see Chapuis and Droz (1958); Langton's introduction 'Artificial life' in Langton (ed.) (1989).
24. Walter (1950), and the follow-up paper 'A machine that learns' (1951).
25. Thus, Grey Walter not only anticipated the notion of autonomous agents, he also observed emergent collective behavior long time before recent work on collective behavior and swarm intelligence (e.g., Varela and Bourgine [eds.] 1992).
26. Grey Walter even thought that 'it would even be technically feasible to build processes of self-repair and of reproduction into these machines' (1950: 45). In this respect he was over-optimistic and did not acknowledge the fundamental problems of 'realizing' biological self-reproduction (compare Kampis 1991). Yet it was Walter who was the first to show that simple control devices could produce lifelike behavior, with learning.
27. Braitenberg (1984) is a classic essay on the synthesis of complex behavior from the interactions of simple components.

28. See Dennett (1987). The intentional stance is the idea that we should not think of our mental vocabulary of 'belief', 'hope', 'fear', etc., as actually standing for genuine mental phenomena, but rather as just a manner of speaking. It is a useful vocabulary for predicting and referring to behavior, but it should not be taken literally as referring to real, intrinsic, subjective, psychological phenomena; it is rather a matter of taking an 'intentional stance' toward any kind of autonomous system (in the intuitive sense of autonomy), whether it be an insect, a robot, or a human being. Even though it clearly differs from J. von Uexküll's certainty about the genuine character of the experiential or subjective content of the Umwelt, both a Dennettian and Uexküllian methodology for the study of behavior require the researcher to take the intentional stance.
29. Or quasi-autonomous might be a better term here. When it comes to real demonstrations (video-taped or 'live' at the conferences) of the various species of situated agents and animats — from the early versions of Braitenberg to the most recent ones — their performance is not too impressive. A typical place to be 'situated' is on a plane floor with smooth obstacles forming plates perpendicular to the floor and no ground ruggedness at all. And yet, these small heroes often get caught in a corner or entangled in the protruding sensors of a companion agent and then, as from heaven sent, the (in)visible hand of their creator comes down and puts them on the wheels again. The art of careful robot education has not got the recognition it deserves.
30. E.g., Weisbuch (ed.) (199l), see also Emmeche (1997) for a review.
31. Cf. Pattie Maes' paper 'Designing autonomous agents', in Maes (ed.) (1990). Brooks' ideas on subsumption architecture were first either ignored or attacked by researchers in traditional robotics, but the approach has been gradually accepted, and in 1991 Brooks received the 'Computers and Thought' award, the highest award in AI. The approach is now known by nicknames such as 'agents design', 'subsumption architecture theory', 'situated agents', 'nouvelle AI', and 'Behavior Based AI'. The term 'autonomous systems', though sometimes designating Brooks's approach too, is often used in a broader sense, including subsumption architecture theory. On the development of Brooks's ideas, see also Levy (1992).
32. An indication of which was when the journal *Robotics* in 1998 changed its name to *Robotics and Autonomous Systems*. See also Levy (1992) and Brooks' papers.
33. Winograd and Flores (1986); Varela et al. (1991); Hendriks-Jansen (1996); Clark (1997).
34. Compare also Andy Clark's comment that 'The similarity between the operational worlds of Herbert [one of Brooks's robots from the 1980s, CE] and the tick [as described by J. von Uexküll] is striking: Both rely on simple cues that are specific to their needs, and both profit not bothering to represent other types of detail' (Clark l997: 25). It remains to be accurately accessed by historians of science to what extent the Umwelt theory really determined the conceptual development within ASR, but Rodney Brooks was clearly influenced (Brooks 1986b). For a more critical use of the Umwelt notion in assessing the merits of ASR, see Sharkey and Ziemke (1998) and their article in this special issue.
35. Maes (1990); Maes does not define the notion of emergence, which seems to be observer-dependent, cf. Cariani (1992); Emmeche (1994a).
36. Compare Steels (1990) who distinguishes between (a) categorial representations and (b) analogical representations, where (a) includes symbolic as well as sub-symbolic (i.e., perceptron-like networks with categories coded in terms of patterns of activation over a collection of units) representations, while (b) includes various types of maps (e.g., for sensory information, a frequency map, a sonar map, a 'smell' map, a color map).

37. A subsumption program is built on a computational substrate that is organized into a series of incremental layers, each (generally) connecting perception to action. The substrate is a network of finite state machines augmented with timing elements. It is best understood in contradistinction to the Good Old Fashioned Robotics paradigm according to which the robot first perceives the environment, then starts to reason about it, tries to build a model of the world and lay plans as how to achieve the goals represented in the robot. Only when this is done, the robot would act, so to speak, by translating its cognition into behavior. Brooks thought that the coupling of action to perception should be more direct, without the 'cognitive bottleneck' of the traditional architectures. This is not to give up rule-following behavior. But the agent should consist of a series of modules (each a finite state machine, hence rule-based, even if the rules may be low-level informational). Information from sensors about the world will be processed according to the rules, but in parallel in each module, and the behavior of the multi-module agent will emerge from the continual series of actions involved. Thus, the subsumption architecture consists of layers of behavior modules that trigger other behaviors when needed. Notice the bottom-up structure: the basic level behaviors cope with objects in the world on a moment-to-moment basis. Low-level behaviors are determined by sensor inputs on the legs, for instance. The next level might be a 'walk' behavior; an even higher one is 'explore' (Brooks 1992).
38. 'Once this commitment [physical grounding] is made, the need for symbolic representations soon fades entirely. The key observation is that the world is its own best model. It is always exactly up to date. It always contains every detail there is to be known. The trick is to sense it appropriately and often enough' (Brooks 1990). Steven Harnad proposed to solve the symbol grounding problem by construction of hybrids of symbolic and non-symbolic sensor-motor systems; close to Brooks' idea of 'physical grounding' (see Harnad 1990).
39. See for instance Peschl (1994) who states that '"representations" can be better characterized as finding a stable relation/covariance between' [the environment] 'and inside the representation/body system. This can be achieved by adaptational/constructive changes in the neural substratum which leads to an embodied dynamics capable of generation functionally fitting behavior ('representations without representations')' (1994: 423).
40. A neglected issue in ASR (see below) is whether an artificial robot really experiences sensations (or anything at all) or have *a body* in the true sense of an organism. From a biological point of view the latter claim is trivially false as a robot and an animal are constructed and maintained by fundamentally different types of processes. Brooks claims real robots (as opposed to computer-simulated robots) to be embodied. As we shall see, this claim is crucial to discuss if we want to assert that they also can have an Umwelt. See also the work of Tom Ziemke and Noel Sharkey.
41. This is Brooks's (1990) argument from time. As evolution of the first simple living organisms on Earth took roughly a billion years, this was a slow process. (Recent evidence questions this estimate and suggests that the appearance of early life was a much faster process). Another billion years passed before photosynthetic plants appeared, and almost one billion and a half years (ca. 550 million years ago) the first vertebrates arrived — to create organisms with information processing systems are rather hard problems. Then things started to move fast; reptiles arrived around 370 million years ago, mammals 250 million years ago, the first primates appeared around 120 million years ago, the predecessors to the great apes just 18 million years ago. Man-like creatures arrived 2.5 million years ago. Man invented agriculture 19,000 years ago, and developed writing and 'expert knowledge' less than 5,000 years ago.

Thus, problem solving behavior, language, expert knowledge, and reason seems to be rather simple, once the essence of being and reacting are available!

42. See comment by Belew (1991), who has worked on machine learning.
43. Stanley Salthe rightly pointed out that the correct relation is not AI = AS, but {AS {AI}}.
44. In fact, a lot of research in artificial life and 'swarm intelligence' has been concerned with understanding the structure of patterns of collective behavior in ants, wasps, and other social insects. See, for example, Deneubourgh et al. (1992), other papers in that volume, and the papers on collective behavior in Morán et al. (1995).
45. *CP* 4.537 fn. 3: 'The type, token and tone are the legisigns, sinsigns and qualisigns discussed in 2.243f'. Compare also 'A quality of feeling, in itself, is no object and is attached to no object. It is a mere tone of consciousness' (*CP* 7.530); and the statement that 'There is a certain tinge or tone of feeling connected with living and being awake, though we cannot attend to it, for want of a background' (*CP* 8.294).
46. This is what Liszka (1996: 46) calls *the inclusion rule*. This rule applies within each of the three major divisions (i.e., according to the feature of the sign itself (qualisign-sinsign-legisign); according to the sign's relation to the object (icon, index, symbol); and according to the sign's power to determine an interpretant (rheme, dicisign, argument)). The true logical implication is that for each possible kind of sign within the ten classes scheme of Peirce, all will include a qualisign (even an argument, which is also a symbolic legisign), even though the qualitative aspect of the sign may not be the dominant aspect. Thus my analysis follows Liszka (and Peirce), but what I give special emphasis is the phenomenal and qualitative aspect of every semiosis.
47. *CP* 2.254. We seldom just experience qualisigns in our Umwelt; they are rather to be thought of as the sensual background of our perception. Merrell (1996: 43) describes the perception of a drawing of a Necker cube as offering an 'example of the of this bare, passive *seeing* in contrast to *seeing as* and *seeing that* such-and-such is the case … seeing the drawing as immediacy entails a feeling or sensing of nothing more than a *quality* (Firstness): Whiteness punctuated with thin intermittent blackness. A split moment later it is *seen* in terms of some *existent* entity ''out there'' in the ''semiotically real'' merely as a set of interconnected lines. But it is not (yet) actively seen as a cube'. Thus the emergence of the cube percept corresponds to the 'development' of a legisign, that includes within it the sinsign and the qualisign; in this sense qualisigns permeate our Umwelt.
48. Though it is indeed possible to reconstruct, externally, a model (in our Umwelt) of the creature's phenomenal world; compare Salthe (this issue) and Cariani (1996).
49. See also Jesper Hoffmeyer's article (this issue) referring to Popper making the same point.
50. A similar remark is briefly stated in a note in a famous essay by Thomas Nagel, where he emphasizes that the subjective character of experience is not analyzable in terms of functional states since these could be ascribed to robots that behaved like people though they experienced nothing: 'Perhaps there could not actually be such robots. Perhaps anything complex enough to behave like a person would have experiences' (Nagel 1974: 392).
51. Interlevel downward causation should not be seen as an instance of the usual (temporal) efficient causation, but rather as a functional and formal cause. See Emmeche, Køppe, and Stjernfelt 2000.

52. I would like to thank Ricardo Gudwin, Jesper Hoffmeyer, Kalevi Kull, Winfried Nöth, Stanley Salthe, and Tom Ziemke for helpful comments and criticism on earlier versions of the article.

References

Anderson, M.; Deely, J.; Krampen, M.; Ransdell, J.; Sebeok, T. A.; and Uexküll, T. von (1984). A semiotic perspective on the sciences: Steps toward a new paradigm. *Semiotica* 52 (1/2), 7–47.

Bateson, G. (1979). *Mind and Nature. A Necessary Unity*. London: Wildwood House. [Used here is the Fontana paperback edition from 1980.]

Beckermann, A.; Flohr, H.; and Kim, J. (eds.) (1992). *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. Berlin: Walter de Gruyter.

Belew, R. K. (1991). Artificial life: A constructive lower bound for artificial intelligence. *IEEE Expert* 6 (1), 8–15, 53–59.

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

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

—(1986b). Achieving artificial intelligence through building robots. Technical Report Memo 899, MIT AI Lab.

—(1990). Elephants don't play chess. In Maes 1990: 3–15.

—(1991a). New approaches to robotics. *Science* 253, 1227–1232.

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

—(1992). Artificial life and real robots. In Varela and Bourgine 1992: 3–10.

Brooks, R. A. and Maes, P. (eds.) (1994). *Artificical Life IV*. Cambridge, MA: MIT Press.

Burks, A. W. (1975). Logic, biology and automata — Some historical reflections. *Int. J. Man-Machine Studies* 7, 297–312.

Cariani, P. (1992). Emergence and artificial life. In Langton et al. 1992: 775–797.

—(1996). Life's journey through the semiosphere. *Semiotica* 120 (3/4), 243–257.

Chapuis, A. and Droz, E. (1958). *Automata: A Historical and Technological Study*, trans. by A. Reid. London: B. T. Batsford.

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

Deneubourg, J-L.; Theraulaz, G.; and Beckers, R. (1992). Swarm-made architectures. In Varela and Bourgine 1992: 123–132.

Dennett, D. C. (1987). *The International Stance*. Cambridge, MA: MIT Press.

Emmeche, C. (1992). Life as an abstract phenomenon: Is artificial life possible? In Varela and Bourgine 1992: 466–474.

—(1994a). *The Garden in the Machine. The Emerging Science of Artificial Life*. Princeton, NJ: Princeton University Press.

—(1994b). The computational notion of life. *Theoria — Segunda Época* 9 (21), 1–30. [Also online at http://www.nbi.dk/~emmeche/cePubl/compnolife.html].

—(1997). Aspects of complexity in life and science. *Philosophica* 59 (1), 41–68. [University of Ghent (actually first appeared in 1999)].

Emmeche, C.; Køppe, S.; and Stjernfelt, F. (2000 [1997]). Levels, emergence and three versions of downward causation. In *Downward Causation*, P. B. Andersen, N. O. Finnemann, P. V. Christiansen, and C. Emmeche (eds.), 3–34. Aarhus: Aarhus University Press.

Gevarter, W. B. (1985). *Intelligent Machines: An Introductory Perspective of Artificial Intelligence and Robotics*. Englewood Cliffs, NJ: Prentice-Hall.
Gregory, R. L. (1981). *Mind in Science. A History of Explanations in Psychology and Physics*. London: Georg Weidenfeld and Nicholson. [Reprinted 1988. New York: Penguin Books].
Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346.
Harrington, A. (1996). *Reenchanted Science. Holism in German Culture from Wilhelm II to Hitler*. Princeton, NJ: Princeton University Press.
Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press.
Hayes, P.; Harnad, S.; Perlis, D.; and Block, N. (1992). Virtual symposium on virtual mind. *Minds and Machines* 2 (3), 217–238.
Hendriks-Jansen, H. (1996). *Catching Ourselves in the Act. Situated Activity, Interactive Emergence, Evolution, and Human Thought*. Cambridge, MA: MIT Press.
Heylighen, F.; Rosseel, E.; and Demeyere, F. (eds.) (1990). *Self-Steering and Cognition in Complex Systems: Toward a New Cybernetics*. New York: Gordon Breach Science Publishers.
Hoffmeyer, J. (1996). *Signs of Meaning in the Universe*. Bloomington: Indiana University Press.
Janlert, L-E. (1987). Modelling change — The frame problem. In Pylyshyn 1987: 1–40.
Kampis, G. (1991). *Self-Modifying Systems in Biology and Cognitive Science: A New Framework for Dynamics, Information and Complexity*. Oxford: Pergamon Press.
Kant, I. (1790 [1951]). *Kritik der Urteilskraft*, trans. by J. H. Bernard (*Critique of Judgment*). New York: Hafner Publishing Company.
Katz, H. and Queiroz, J. (1999). Notes about representation (in the Wild) & Continuum. Paper presented at the "2° Seminário Avançado de Comunição e Sémiotica: Novos Modelos de Representação: vida artificial e inteligência artificial", August 18–20, 1999, São Paulo, Brazil.
Langton, C. G. (ed.) (1989). *Artificial Life* (=Santa Fe Institute Studies in the Sciences of Complexity 6). Redwood City, CA: Addison-Wesley.
Langton, C. G.; Taylor, C.; Doyne Farmer, J.; and Rasmussen, S. (eds.) (1992). *Artificial Life II* (=Santa Fe Institute Studies in the Sciences of Complexity 10). Redwood City, CA: Addison-Wesley.
Lemke, J. (2000). Opening up closure: Semiotics across scales. In *Closure: Emergent Organizations and Their Dynamics* (=Annals of the New York Academy of Sciences 901), Gertrudis van de Vijver and Jerry Chandler (eds.), 100–111. New York: New York Academy of Sciences.
Levy, S. (1992). *Artificial Life. The Quest for a New Creation*. New York: Pantheon Books.
Liszka, J. J. (1996). *A General Introduction to the Semeiotic of Charles Sanders Peirce*. Bloomington: Indiana University Press.
Lilienfeld, R. (1978). *The Rise of Systems Theory*. New York: John Wiley & Sons.
Maes, P. (ed.) (1990). *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*. Cambridge, MA: MIT Press.
Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition*. Dordrecht: Reidel.
Mayr, O. (1970). *The Origins of Feedback Control*. Cambridge, MA: MIT Press.
Mele, A. R. (1995). *Autonomous Agents: From Self-Control to Autonomy*. Oxford: Oxford University Press.
Merrell, F. (1996). *Signs Grow: Semiosis and Life Processes*. Toronto: University of Toronto Press.
Meyer, J-A. and Guillot A. (1991). Simulation of adaptive behavior in animats: Review and prospects. In *From Animals to Animats*, J-A. Meyer and S. Wilson (eds.), 2–14. Cambridge, MA: MIT Press.

Meystel, A. (1998). Multiresolutional Umwelt: Toward a semiotics of neurocontrol. *Semiotica* 120 (2/3), 243–380.
Morán, F.; Moreno, A.; Merelo, J. J.; and Chacón, P. (eds.) (1995). *Advances in Artificial Life*. Berlin: Springer.
Moreno, A.; Ibañez, J.; and Umerez, J. (1997). Cognition and life: The autonomy of cognition. *Brain & Cognition* 34 (1), 107–129.
Nagel, T. (1974). What is it like to be a bat. *The Philosophical Review* 83 (October), 435–450. [Reprinted 1981 in *The Mind's I,* D. R. Hofstadter and D. Dennett (eds.), 391–403. New York: Penguin Books.]
—(1986). *The View from Nowhere*. Oxford: Oxford University Press.
Needham, J. (1943). *Time: The Refreshing River*. London: George Allen & Unwin.
Newell, A. (1980). Physical symbol systems. *Cognitive Science* 4, 135–183.
Nöth, W. (1997). Representation in semiotics and in computer science. *Semiotica* 115 (3/4), 203–213.
Pattee, H. H. (1989). Simulations, realizations, and theories of life. In Langton 1989: 63–77.
Peirce, C. S. (1931–1958). *Collected Papers of Charles Sanders Peirce*, 8 vols., Charles Hartshorne, Paul Weiss, and Arthur Burks (eds.). Cambridge, MA: Harvard University Press. [Reference to Peirce's papers will be designated *CP*.]
—(1955). *Philosophical Writings of Peirce*, ed. by Justus Buchler. New York: Dover Publications.
Peschl, M. F. (1994). Autonomy vs. environmental dependency in neural knowledge representation. In Brooks and Maes 1994: 417–423.
Pylyshyn, Z. W. (ed.) (1987). *The Robot's Dilemma. The Frame Problem in Artificial Intelligence*. Norwood, NJ: Ablex Publishing.
Rasmussen, S. (1992). Aspects of information, life, reality and physics. In Langton et al. 1992: 767–773.
Richards, R. J. (1987). *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago: University of Chicago Press.
Salthe, S. N. (1988). Notes toward a formal history of the levels concept. In *Evolution of Social Behavior and Integrative Levels* (=The T. C. Schneirla Conference Series 3), G. Greenberg and E. Tobach (eds.), 53–64. Hillsdale, NJ: Lawrence Erlbaum.
Santaella Braga, L. (1994). Peirce's broad concept of mind. *European Journal for Semiotic Studies* 6 (3/4), 399–411.
Sharkey, N. E. and Ziemke, T. (1998). A consideration of the biological and psychological foundation of autonomous robotics. *Connection Science* 10 (3/4), 361–391.
Searle, John (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
Sebeok, T. A. (1991). *A Sign is Just a Sign*. Bloomington: Indiana University Press.
Sheets-Johnston, M. (1998). Consciousness: A natural history. *Journal of Consciousness Studies* 5 (3), 260–294.
—(2000). The formal nature of emergent organization. In *Closure: Emergent Organizations and Their Dynamics* (=Annals of the New York Academy of Sciences 901), Gertrudis van de Vijver and Jerry Chandler (eds.), 320–331. New York: New York Academy of Sciences.
Steels, L. (1990). Exploiting analogical representations. *Robotics and Autonomous Systems* 6, 71–88.
Uexküll, Jakob von (1940). *Bedeutungslehre* (=Bios 10). Leipzig: Johann Ambrosius Barth. [English trans. by Barry Stone and Herbert Weiner as *The Theory of Meaning*, ed. with a Glossary by Thure von Uexküll. *Semiotica* 42 (1), 25–87.]
Uexküll, Thure von (1982a). Introduction: Meaning and science in Jakob von Uexküll's concept of biology. *Semiotica* 42 (1), 1–24.

—(1982b). Semiotics and medicine. *Semiotica* 38 (3/4), 205–215.
—(1986a). From index to icon, a semiotic attempt at interpreting Piaget's developmental theory. In *Iconicity. Essays on the Nature of Culture. Festschrift for Thomas A. Sebeok on his 65th birthday*, Paul Bouissac, M. Herzfeld, and R. Posner (eds.), 119–140. Tübingen: Stauffenburg Verlag.
—(1986b). Medicine and semiotics. *Semiotica* 61 (3/4), 201–217.
—(1989). Jakob von Uexküll's Umwelt-Theory. In *The Semiotic Web 1988*, T. A. Sebeok and J. Umiker-Sebeok (eds.), 129–158. Berlin: Mouton de Gruyter.
Uexküll, Thure von; Geigges, W.; and Hermann, J. M. (1993). Endosemiosis. *Semiotica* 96 (1/2), 5–51.
Van de Vijver, G.; Salthe, S. N.; and Delpos, M. (eds.) (1998). *Evolutionary Systems. Biological and Epistemological Perspectives on Selection and Self-Organization*. Dordrecht: Kluwer.
Varela, F. J. and Bourgine, P. (eds.) (1992). *Toward a Practice of Autonomous Systems. Proceedings of the First European Conference on Artificial Life* [ECAL91]. Cambridge, MA: MIT Press.
Varela, F. J.; Thompson, E.; and Rosch, E. (1991). *The Embodied Mind. Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.
Walter, W. G. (1950). An imitation of life. *Scientific American* 182 (5), 42–45.
—(1951). A machine that learns. *Scientific American* 185 (2), 60–63.
Weisbuch, Gérard (ed.) (1991). *Complex System Dynamics* (=Santa Fe Studies in the Sciences of Complexity Lecture Notes 2). Redwood City. CA: Addison-Wesley.
Wiener, N. (1948). Cybernetics — Or Control and Communication in the Animal and the Machine. Cambridge, MA: MIT Press.
Wilson, S. W. (1991). The animat path to AI. In *From Animals to Animats*, J-A. Meyer and S. Wilson (eds.), 15–21. Cambridge, MA: MIT Press.
Winograd, T. and Flores, F. (1986). *Understanding Computers and Cognition: A New Foundation for Design*. Norwood, NJ: Ablex Publishing.
Ziemke, T. and Sharkey, N. E. (eds.) (1998). Biorobotics. *Connection Science* 10 (3/4). [Special issue.]

Claus Emmeche (b. 1956) is Associate Professor at the Niels Bohr Institute at the University of Copenhagen ⟨emmeche@nbi.dk⟩. His research interests include biosemiotics, theoretical biology, artificial life, and the philosophy of science. His major publications include *The Garden in the Machine* (1994), 'The computational notion of life' (1994), 'Aspects of complexity in life and science' (1997), and 'The Sarkar challenge to biosemiotics: Is there any information in a cell?' (1999).