Epistemic Autonomy in Models of Living Systems

Erich Prem
The Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Wien, Austria
erich@ai.univie.ac.at

Abstract

This paper discusses epistemological consequences of embodied AI for Artificial Life models. The importance of robotic systems for ALife lies in the fact that they are not purely formal models and thus have to address issues of semantic adaptation and epistemic autonomy, which means the system's own ability to decide upon the validity of measurements. Epistemic autonomy in artificial systems is a difficult problem that poses foundational questions. The proposal is to concentrate on biological transformations of epistemological questions that have lead to the development of modern ethology. Such an approach has proven to be useful in the design of control systems for behavior-based robots. It leads to a better understanding of modern ontological conceptions as well as a reacknowledgement of finality in the description and design of autonomous systems.

Key words: epistemic autonomy, embodied AI, epistemology, robotics, theoretical biology, finality, teleology, ontology.

1 Introduction

1.1 Artificial Life

Post-modern epistemologists and commentators of science usually regard interdisciplinarity as a distinguishing advantage of the newly founded 20th century disciplines. However, interdisciplinarity does not come without a price to pay and the field of Artificial Life (ALife) provides a good example for this claim. The many classical disciplines in ALife that contribute to its prosperity are not necessarily striving for a common goal. While the computer scientists are interested in discrete dynamical and self-modifying systems, the chemists may study self-steering processes of catalytic reactions of enzymes. Again others, typically biologists, try to find answers to the ultimate questions of life: its origins, its course, its end.

Except for the few philosophers, who usually tend to stay outside the field they comment on, scientists in ALife can be easier classified as such based on the methods they are using. These methods have been described as [Langton 89]

- the synthesis of life-like behaviors within artificial systems,
- the study of ongoing dynamics rather than a final result,
- and the construction of systems which exhibit emergent phenomena.

Therefore, what unifies researchers in ALife are their methods rather than their goals. Of course, there is a general interest in "life" as the phenomenon under study. Unfortunately though, life is too vague to ensure a common direction of research. Evidence for this claim can be found in the fact that some are interested in "systems that exhibit phenomena of living systems", others search for the origins of chemical reproduction, again others try to solve the philosophical problems of auto-poiesis. And a completely different set of researchers, namely roboticists, try to construct physical systems which exhibit some behavioral similarities with that of living animals.

This paper aims at a better understanding of differences in the research aims and methods within ALife. It will be argued that the study of autonomous robotic systems is profoundly different from purely formal approaches to the study of ALife. The construction and analysis of robotic systems is more in accordance with the aims of theoretical biology. It can provide us with a better understanding of such debatable concepts as finality, anticipation, representation and teleology. This understanding is based on the emphasis of the animal's view of the world and its ontological perspective.

In the following section, we take a closer look at purely formal approaches within ALife.

1.2 Life as a formal property

Many proponents in the field of ALife regard "life" as a property of the *oganization* of matter rather than a material phenomenon. Some have even argued that the "material" that realizes life is irrelevant to the study of its properties. Others, mainly epistemologically interested researchers, have argued that "Life is matter with meaning." [Pattee 95] and that living systems are material structures with memory by virtue of which they construct, control and adapt to their environment. The fact that our notion of living systems stems from biological realizations which are physical, i.e. material systems, is obvious. The development of a purely formal account of the phenomenon of life is, historically, paralleled by formal accounts of physical phenomena, of cognition and of intelligence.

The usual argument that is used to support such a methodological approach is based on the fact that the material underlying living phenomena can take on a wide range from single cells to elephants. Matter in these living systems, however, is of a highly individual nature. In fact, it means individuality per se. But a scientist interested in "life" as a general property seeks to describe what is common to all living systems. A clear account of what it means to be alive is one of the basic, still unresolved problems of biology, maybe best described by [Schrödinger 67]. Instead of an explanation that could give a complete description of the "essence" of living systems, of "what it means to be this" (τὸ τί ἦν εἶναι), computationalists are satisfied with systems that exhibit properties which living systems (or intelligent, or cognitive) also possess. This restriction to fomalized properties alone happens through a process of abstraction, in which an I-O behavior of the system interface turns into the center of scientific interest. It has been argued often before that it is generally impossible to reason backwards from the I-O behavior of a system to the system that realizes a function [Rosen 85, Pattee 95, Prem 95a]. Nevertheless, there still is a strong fascination that emanates from the formal dating back to Aristotle.

In the first book of his "physics" Aristotle develops his theory of first principles, most importantly of form and matter. The argument starts with a discussion of the origin of movement and change, for Aristotle the starting point for all scientific explanations. Change is identified by means of some substratum that is able to manifest the change. This substratum can be easily identified with matter. The ability of coming to be something, however, must be based on a specific form of privation. For Aristotle, "pure" matter shows this kind of privation due to its lack of *form*. Formless matter does not identify things properly and cannot explain change completely. Formlessness is a positive deficiency of matter. Although matter is the first substratum of everything, from which something comes, it does not have an ontic or epistemic status of its own. Matter "is" only (or rather, "is only knowable") because of its lack of, and therefore, potential for form. The substratum, in a next step, is not only form-able, it *needs* to be formed. This results in the primacy of form over matter. Knowledge of things must be based on formal principles rather than on material substrates, as the following quote from Aristotle's Metaphysics shows:

> By form I mean the essence of each thing and its primary substance. For even contraries have in a sense the same form; for the substance of a privation is the opposite of substance, e.g. health is the substance of disease... [Aristotle, 7,1032b 1]

Additionally, "form" is one of the four original causes that explain the why of things (formal, material, efficient, and final cause).

> In one of [the four senses] we mean the substance, i.e. the essence (for the 'why' is reducible finally to the definition, and the ultimate 'why' is a cause and principle); in another the matter or substratum, in a third the source of the change, and in a fourth the cause opposed to this, the purpose and the good (for this is the end of all generation and change). [Aristotle, 1, 983b 25].

The success of Aristotle's paradigm of form in ALife is based on the coincidence that computers are excellent in reproducing form. Unfortunately though, computers cannot reproduce matter. This is why a major branch of ALife is busy with the construction of computer simulations of metaphorical physics or chemistry. In fact, as [Rosen 91] has pointed out, computers implement formalisms in perfect analogy to the first three causes of Aristotle. Axioms of a formalisms may be considered material causes, efficient causes can mean production rules and formal causes can be identified with a particular sequence of production rules. Finality is usually omitted, because it does not respect the flow of "formal time", indeed it appears to violate this flow because the cause is later than the effect.

I have argued before that there is evidence that ALife should not be pursued as such a purely formal discipline [Prem 95a]. Instead, there exists a small subfield of ALife that has the potential of contributing to an understanding of living systems on a basis which is far from being only formal.

2 Embodied Artificial Intelligence

2.1 A departure from formal models

In its short history, embodied Artificial Intelligence has challenged a sizeable number of foes. Among the list of opponents we find classical robotics and Artificial Intelligence (AI) in computer science, cognitivism in psychology, and objectivism in philosophy. The provocation lies in embodied AI's attack on a fundamental assumption of modern Western science. Dating back to medieval philosophy (or to Descartes, if you prefer) this assumption has been the primacy of the mental in the study of human cognition. 'Mental' here does not only refer to the opposite of 'physical' but also means 'rational' which is often considered opposed to emotion and intuition. To the extent that embodied AI tries to replace this predominance of the mental and rational by an emphasized acknowledgement of the bodily basis of cognition [Brooks 91] it threatens the disciplines mentioned above, which have a long tradition in disregarding the human body. In the context of ALife, embodied AI also threatens ALife's appraisal of the formal. Typical research in embodied AI proceeds by constructing physical robots with real-world dynamics. The dynamic behaviors of these robots are studied and generated by complex control systems and high interaction rates of the system and its environment.

It is easy to see that embodied AI proceeds quite differently from conventional robotic research and Artificial Intelligence [Prem 97a, Prem 97b]. Embodied AI has been shown to improve on the dynamical qualities of intelligent embodied systems, i.e. to get the interaction dynamics right. However, the structures of the new embodied control programs are very different from the traditional ones. Robot control tasks are no longer considered to be of a purely formal nature. To the roboticist in embodied AI it is unimportant whether she realizes a function by means of a computational procedure or a physical characteristic of the robot. Getting the interaction dynamics right is more important. Thus the control task is not considered as a purely computational I-O mapping, but as a combination of a physically and informationally transducing process that serves to generate "intelligent" bodily behavior. The realization of these control tasks do not happen by means of functional modules with clearly defined interfaces. It is achieved by dynamically interacting processes that are tightly coupled with the environment and with environmental time rather than with internal state-transition time (cf. [Prem 97b]). The system-environment coupling of robots is based on effectors and measurement devices. The use of meters marks the departure from purely formal models most clearly.

2.2 Measurement: semantic adaptation and epistemic autonomy

Many critics of formal ALife approaches have concentrated on the semantic gap between simulation and measurement [Rosen 91, Pattee 92, Pattee 95]. It was already John von Neumann who pointed out that results of measurements, choices of observables, the construction of measurement devices and the measurement process itself cannot, in principle, be formalized [von Neumann 55, von Neumann 66]. The reason lies in the fact that the process of measurement is not of a purely formal nature. Measurement is a process in which two dynamical systems interact. It is true that the interaction can be interpreted as a mapping of complex situation to simple patterns, but this view disregards the inherently dynamic nature of the measurement process [Rosen 78]. There are two main problems with measurement for roboticists: One is the *construction* of measurement devices that optimally support the robot. The other is the *interpretation* of numbers delivered by some existing meter.

The construction of meters in biological systems is a process of semantic adaptation in which new observables are developed by the system. Semantic adaptation cannot be simulated [Cariani 90] and is very hard to be reproduced artificially [Pask 58, Pask 61]. Robot engineers cannot wait for evolution and must design meters that support the dynamical robot behavior. Embodied AI has developed a specific strategy for the design of these meters that can be sharply distinguished from previous approaches. Instead of developing complex sensors with complex interpretation routines, the emphasis is now on quick sensor interpretation. This has lead to an increased use of simple sensors (switches, IR, etc.) with higher rates of sensor readings to support the system dynamics.

These high rates of sensor interpretation are possible, because the meaning of the sensor reading is *physically* highly restricted. Instead of using a camera and searching for constraints on video data that allow the identification of a certain object, robots are now equipped with physically constrained object recognition sensors, e.g. a special "soda-can sensor" [Connell 90]. This strategy emphasizes the physical design and coupling of a system at the cost of formal sophistication (at least with respect to sensor readings). From an epistemological point of view, the truth about the predicate "soda-can in front of gripper" is materialized in physical interaction (especially if we take into account that the robot can also check whether the can is empty by trying to lift it) up. This marks a clear departure from previous approaches in which truth conditions were constructed as formal constraints on streams of numbers.

The deeper reason for this strategy lies in the necessity to equip embodied systems with *epistemic autonomy.* A robot must be able to find out whether its sensor readings are distorted and, even more importantly, exactly when a measurement has occured. The correct *interpretation* of sensor readings is of vital importance for the generation of useful system actions. This epistemic condition on autonomous systems is central to all ALife models, indeed to all living systems. It arises in autonomous embodied models, because no humans are available to interpret the data and a pre-selection of valid data is impossible for practical reasons. (This is in sharp distinction from purely formal domains, in which the interpretation of artificial sensors can, at least in principle, always be reduced to formal constraints. The ability for whiskers to break, however, is usually not modeled in any robot simulation.)

While semantic adaptation has been recognized as important for living systems and as very problematic for ALife, epistemic autonomy has not found similar attention so far. (The general importance of meter interpretation has been addressed by many authors, e.g. in the work of H.H. Pattee and R. Rosen. The importance of "system detectable error" has been argued by [Bickhard & Terveen 95, Bickhard 97] from an embodied perspective and is discussed in Sec. 4.2.) It should be emphasized that epistemic autonomy is a phenomenon at the level of the individual, whereas semantic adaptation is usually regarded as an evolutionary phenomenon. Nevertheless, epistemic autonomy has been recognized as important in embodied AI due to the urging need to construct systems that can decide upon the validity of their sensor readings. Let us now take a closer look on the epistemic consequences of such a perspective in relation to theoretical biology.

3 Theoretical biology and functional circuits

I propose a concentration on a view of system epistemics that is more oriented towards biology and has a sensory-motor perspective rather than a merely formal or even evolutionary one. In such a perspective the fundamental building blocks of epistemic conditions are control schemata for motor patterns that arise from perceptual interaction with the system environment. The drives for the system arise from within the system as needs or goals.

As soon as 1928 Jakob von Uexküll described a view of biology which bases the study of animals on the animal's view of the world rather than on a scientist's "objective" view of the animal and its environment. This is basically a Kantian turn in producing better predictions of how an animal will behave in a given context.

As an example consider the difference between the two following descriptions of the tick's feeding behavior:

1. The tick attacks warm-blooded animals like humans or deer when they make contact with the trees or grass inhabited by the tick.

2. The tick bites when making contact with anything which has a superficial temperature of 37° C and emits butyric acid.

While the first description is immediately easy to understand, the second certainly has a higher predictive value. The analysis which is necessary to come up with the second way of describing the tick behavior consists in a careful study of a tick's sensory organs and reflexes. In fact, the second version is more a description of *how the tick sees the world* in human terms. For the tick there are no humans, deer, trees, grass, etc. All that governs the tick behavior in the feeding context are specific features of two environmental qualities: temperature and chemical concentration.

> The sensations of the mind become properties of things during the construction of the world, or, one could also say, the subjective qualities construct the objective world. [J.v.Uexküll]

However, at the point where Kant's considerations lead to a discussion of categories as the final set of tools of reason to bring the "manifoldness of experience into the unity of concepts", von Uexküll develops descriptions of sensor (and actuator) spaces. His intention is to describe, how the

> marking signs of our attention turn into marks of the world. [Ibd.]

The source of this transformation process is formed by goal-driven interaction with the environment. The basic construct for explaining this interaction space is the description of *functional circuits.* Figure 1 depicts Uexküll's view of such a circuit (slightly adapted here). A "thing" in the animal's world is only "effector-" and "receptor-bearer". It can be thought of as a generator for signals to receptor organs and as a receptor of manipulations through effectors.

The formation of sensory experience is not only based on inter-*action.* Even more importantly, the interaction has a specific purpose. Such a purpose turns the encountered object from a collection of merely causally operating parts of physical entities into a meaningful assembly of things which are integrated in a purposeful whole. The essential point is *to understand how the thing is embedded in an action and how this action is embedded in a purposeful interaction with the world.* In order to fully understand the system's world, our task consists in the dissection of the functional world (i.e. the whole of the subject's functional circuits).

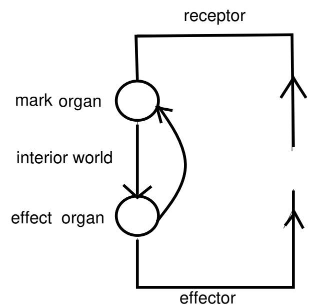

Figure 1: Action circuit as described by Jakob von Uexküll (slightly modified).

Such a point of view is surprisingly close to the credo of embodied AI where the descriptive strategy outlined above is turned into a design method. Starting from functional interaction circuits, the engineer tries to develop a minimalist architecture that fulfills the system requirements. An example for this strategy can be found in [Connell 90].

Summarizing Uexküll's position, there can be no understanding of animals without clarifying how they see the world, or better, what makes up the animal's world. Most notably, no such understanding seems possible without having gained insight into the animal's meaningful whole of functional circuits. To the modern, enlightened scientist such a view is dangerously close to the teleology which has been systematically eliminated from biology in the last century. However, there is a perfectly scientific version of finality that can help in the explanation and construction of embodied AI systems. Such a turn in describing representational elements in embodied AI systems even seems necessary, as will be argued in the next section.

4 Finality

4.1 Anticipation

Consider an adaptive autonomous system that exhibits physical interaction with its environment. A part of such a behavioral system [Brooks 85] is schematically depicted in figure 2.

In this system, the behavior generated by the transfer function is learned based on a training signal. Let us assume that the training signal serves to optimize some criterion that is of importance to the system. It might, for example, assist in the provision of food. Following a description by [Rosen 85], we realize that the system's input is I, while the adaptation is determined by the optimization criterion r. Of course, it is reasonable to believe that there is a linkage between the "predicate" to be learned and the observables of the

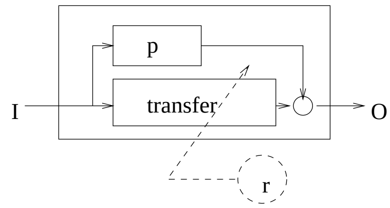

Figure 2: Behavioral module (after [Connell 90]) that transfers an input to an output if the applicability predicate p is true. The transfer function is learned based on some training signal r.

system environment I. Two things happen in this picture.

Firstly, the learning mechanism that selects the proper parameters generates a picture of the external linkage (between I and the predicate to be learned) *within* the system. Secondly, the adaptation must on principle be in a certain sense slower than the system-environment interaction. Thus, the system implicitly generates a *model* of the linkage, and also, of the system-environment interaction. (For an extensive treatment of these system-theoretic properties cf. [Rosen 85].)

The result of such a learning or selection mechanism is a transfer function that "predicts" external reinforcement, i.e. it drives the system in a way that fitness is optimized before it is evaluated. This is why Rosen calls such systems *"anticipatory."* Representations (data structures, models) in the transfer functions become shaped based on their predictive value with respect to maximizing fitness. As a further consequence, these representations must be properly explained with reference to the future outcome of the system's interaction with its environment. This results in a finalistic or teleological terminology. (A discussion of this process based on a computational model of symbol grounding can be found in [Prem 95b].)

Note that the selection mechanism itself works perfectly causally. It operates based on inputs and "rewards". But the generation of some kind of internal model (be it symbolic, connectionist, or statistical) makes it necessary to change the merely causal description of the system and the system's "representational" framework to either a finalistic or, probably, intentional one. The physical embodiment of our system is an important fact in this context. Bodily interaction is hard to describe and measure perfectly. Therefore, the subjective component becomes so strong that objectivity lies in the system's "objectives" rather than in purely physical properties. (Without this being unphysical, of course.)

4.2 Epistemic Autonomy and Representation

As innocent as this descriptional framework may look, it has a rather strong influence on the system's representational framework. It is now likely that sensory impressions of the system are categorized in classes that form items of the same usefulness to the system. In the same sense that "chairs" are properly described by their function "for sitting" for humans, objects in an embodied system's environment will now be classified due to their functional properties. It is clear that such a representational frame can be conceptually opaque in relation to human concepts.[1]

Additionally, the system will appear to behave depending on future events. This happens, because the actions are chosen so as to maximize reward or fitness that is evaluated later, based on an internalized goal-oriented model of system-environment interaction. This model, however, is based on the system's past experience.

There is evidence at the neurophysiological level that exactly this kind of finalistic indicative representation plays a major role in sensory-motor body-environment interaction [Tanji et al. 94]. The subject centered viewpoint of Uexküll has also been well supported by neurophysiological evidence. Experiments by [Graziano et al. 94] show that premotor neurons play a major role in the coding of visual space. The evidence suggests that the encoding of the spatial location of an object happens in arm-centered coordinates rather than using a retinocentric represenation. This again points to the way how system-environment interaction of an embodied system creates models that are heavily oriented by the system's functional needs.

Epistemic autonomy in an adaptive, autonomous, embodied agent is based on the kind of interactive representation outlined above. [Bickhard 97] makes the case for this kind of representation in order to allow for "system detectable representational error". As Bickhard rightly points out, this kind of representation is necessary for action selection in agents that operate in insufficiently predictable environments. The detection of representational error in such schemes is based on internal interaction outcomes. The system must anticipate internal system states that indicate the success of the associated interaction. The representational content of such schemes consists in presuppositions about the environment. The indication implicitly stands for those poperties of the environment, in which the indicated outcomes occur. Thus the representation can be false, when the anticipated outcome does not occur.

Bickhard's "system detectable representational error" also contributes to a better understanding of epistemic autonomy. The results of measurements of the environment in the model presented above amount to anticipations of outcomes of interaction. In this view, measurements are elements of an anticipatory process themselves, because they map a multiplicity of events on few states which are indicative for successful interaction.

Of course, this kind of epistemic autonomy does not mean that all problems for the animal are solved. It is not automatically possible for the individual

[1] "And if a lion could speak, we would not understand it." [Wittgenstein 53]

to tell which of many subsystems is in error when the indicated (anticipated) outcome does not occur. It can still be the interaction that failed, an ungrounded anticipation, or a wrong meter reading. In practice, such a clarification must happen by means of comparisons with other meters and other experiences of the autonomous system.

5 Ontological Aspects

This finalistic view brings with it the development of a rather distinct system ontology. The ontological position described here is so surprisingly similar to the existential-ontological philosophy of Martin Heidegger [Heidegger 27] that it is worth describing a few points of contact between both ontologies. Our notion of *things* in the animal's world can be best compared to what Heidegger calls *equipment*. In the human Being's everyday practices things in our world make sense because we can use them.

> We shall call those entities which we encounter in concern *"equipment"*. In our dealings we come across equipment for writing, sewing, working, transportation, measurement. The kind of being which equipment possesses must be exhibited. [Heidegger 27, p.68, taken from [Dreyfus 90]]

The entities that will be encountered this way are not objects in the above sense. We do not simply add a functional predicate to them. *Dealing* with them is our primordial way of having them, not some bare perceptual cognition. To paraphrase Heidegger, "hammering" does not know about this property of being a tool. Instead, the more we are immediately engaged in coping with the problem of fixing something, the less the hammer is taken as an object which can be used in-order-to hammer [Heidegger 27, p.69]. Strictly speaking, for Heidegger nothing like one equipment in this sense exists. This is because anything which we are using is embedded in a whole of multiple references to other tools and purposes. The hammer thereby refers to nails, tables, wood, etc.–i.e. a whole world of equipment and also of meaningful coping with the world. As long as we are engaged in "hammering"–in a purposeful dealing with equipment–and this equipment simply is "available", we do not even think about it. In such a situation the tools are simply "ready-at-hand".

> The world presents itself in the equipmental nexus, in the reference to a previously seen whole. [Heidegger 27, p.75, my translation]

The world does not consist of things which are "ready-at-hand", because it is only in situations of breakdown that the equipment can be recognized as one thing primarily identified by its sensory or physical properties. In these situations the things are deprived of being "ready-at-hand", creating mere *occurrentness*.

For Heidegger then, the fact that the world usually does not present itself as a world (in the usual scientific sense of the word) is the

> condition of the possibility of the non-entering of the available from the inconspicuous phenomenal structure of this being-in-itself. [Heidegger 27, p.75, my translation]

This view opposes any tradition which believes that things can be identified with reference to their sensory properties. Basically, this belief is based on the Cartesian assumption that extension must be essential characteristic of substance.

> [...] *Descartes* is not merely giving an ontological misconception of the world, but that his interpretation and its basis have lead to *skipping* the phenomenon of the world as well as the being of the [...] innerworldy being. [Heidegger 27, p.95, my translation]

In the end, this is one of the main sources of the problems of traditional robotics. From the idea that sensory and physical properties would be primordial it follows that a physical theory must be used to decide upon (detect, describe, deal with) objects encountered in the world. Moreover, such a theoretical approach must be used to find out whether a table could also be used as a chair. Any usage of tools and any way of dealing with the world therefore have to be explained with respect to those sensory qualities. In a (remotely) existential-ontological view, however, this problem simply does not arise in this way, because dealing with things for a specific purpose is the prevailing mode of encountering them, or rather: to create them. The argument therefore, is not that theoretical objects cannot exist, but that their functional properties must remain inaccessible if functions are not taken as the primordial source of creating things.

Contrary to what people in the field of traditional AI have proposed (perhaps most prominently [Minsky 85]), "functions" may not be some additional property attached to an object, but at the very heart of what things actually are, i.e. of what there is in the world. The conditions of the possibility of object constitution are, of course, constrained by the sensory system. Knowledge about the nature of objects can only be gained by understanding the different actions of the system. The actions, and the related behaviors, must be based on understanding functional circuits. For the system engineer this means that the primary task consists in the design of a functional world of the autonomous system. Such a system, hence, will never be *auto*-nomous, but only *hetero*-nomous. For the ALife scientist, this implies an increased interest in the analysis of functional circuits and in the circular causation of measurements, representations, and action selection.

Heidegger's approach can also be sharply contrasted with Aristotle's ontological primacy of the formal. In his work on technology he explicitly mentions that Socrates and Plato think of essence as something persistent [Heidegger 54]. This persistence is found in the formal appearance of things ($\epsilon\tilde{\iota}\delta o\sigma$), what Aristotle later called ($\tau\grave{o}\ \tau\acute{\iota}\ \tilde{\eta}\nu\ \epsilon\tilde{\iota}\nu\alpha\iota$). As opposed to this view, Heidegger argues for a primacy of effect and work[2] (in the sense of "being at work"). The essence of things therefore lies in their potential or real effects on others, in the induction of changes. Such a view is in close proximity to the practical experiences of roboticists. The first task during the construction of a robot always consists in studying the mutual effects that environment and robot will or must have on each other.

Heideggerian ontology is radically different from the conventional physical view. In my interpretation[3] it acknowledges a teleological element in nature, however, without explicitly mentioning it. This fact is reminiscent of the account that early biologists have given of finality and "entelechie" as well as of Aristotle's discussion of final causation. But in contrast to these accounts, today we need not recur to any kind of divine authority. Instead, the kind of anticipatory adaptation in any autonomous system that physically interacts with the world and has been described in this paper, leads to a most natural account of these teleological ontologies.

6 Conclusion

Embodied Artificial Intelligence is a field of research that can productively contribute to ALife problems in a way that is very different from comparable approaches. The acknowledgement of real physics in embodied AI questions the validity of purely formal accounts of living phenomena on several distinct levels. One of these phenomena is the crossing of the epistemic boundary that lies between a system and its environment. In real life, this boundary is overcome by measurement devices, which introduces the need for epistemic autonomy in such systems. Truth conditions on sensory data are not of a purely formal nature due to the reference to physics and system-environment interaction. Measurement and adaptation in an autonomous embodied system automatically means the development of representation schemes which are based on anticipation and interaction outcome. An understanding of these representations and how they come about lies at the very heart of theoretical biology.

Such an account of autonomy also lends itself nicely to a better understanding of modern ontologies and thus has the potential for changing our understanding of knowledge and the way we humans think about ourselves. Most importantly, an emphasis of embodied AI models of living systems ensures a reacknowledgement of natural elements, be they evolution, biology, ethology, or physiology.

[2] Originally Heidegger uses the verb "wesen", which is derived from the noun "Wesen", which means "essence". Interestingly, the German word for "creature" is also "Wesen".

[3] in fact, it is the Wittgensteinian interpretation of [Dreyfus 90]

7 Acknowledgments

The Austrian Research Institute for Artificial Intelligence is sponsored by the Austrian Federal Ministry of Science and Transport. The author gratefully acknowledges support from Rodney Brooks and the MIT AI Lab.

References

[Aristotle] Aristotle. Metaphysics.

[Bickhard & Terveen 95] Bickhard M.H., Terveen L. 1995. *Foundational Issues in Artificial Intelligence and Cognitive Science.* Elsevier Science Publishers.

[Bickhard 97] Bickhard M.H. 1997. The Emergence of Representation in Autonomous Agents, in Prem E. (ed.) *Epistemological Issues of Embodied AI.* Spec. Iss. of Cybernetics & Systems, to appear.

[Brooks 85] Brooks R.A. 1985. A Robust Layered Control System for a Mobile Robot. *AI-Memo 864.* Cambridge, MA: AI-Laboratory, Massachusetts Institute of Technology.

[Brooks 91] Brooks R.A. 1991. Intelligence without Representation. In Special Volume: Foundations of Artificial Intelligence, *Artificial Intelligence,* 47(1–3).

[Cariani 90] Cariani P. 1990. Implications from Structural Evolution: Semantic Adaptation, in Caudill M.(ed.), *Proceedings of the International Joint Conference on Neural Networks.* Hillsdale, NJ: Lawrence Erlbaum, pp. 47–50.

[Connell 90] Connell J.H. 1990. *Minimalist Mobile Robotics.* San Diego, C.: Academic Press.

[Dreyfus 90] Dreyfus H.L. 1990. *Being-in-the-world.* Cambridge, MA.: MIT Press.

[Graziano et al. 94] Graziano M.S.A., Yap G.S., and Gross C.G. 1994. Coding of Visual Space by Premotor Neurons. *Science,* 11 November 1994, 266, pp. 1054–1057.

[Heidegger 27] Heidegger M. 1927. *Sein und Zeit.* (Being and Time.) Tübingen: Niemayer.

[Heidegger 54] Heidegger M. 1954. Die Frage nach der Technik. (The question for technology.) In M. Heidegger *Vorträge und Aufsätze.* Pfullingen: Günther Neske, pp. 9–40, (6th ed. 1990).

[Kant] Kant, I. 1781. *Kritik der reinen Vernunft.* (A critique of pure reason.)

[Langton 89] Langton C.G. (ed.) 1989. *Artificial Life.* Reading, MA: Addison-Wesley.

[von Neumann 55] von Neumann J. 1955. *Mathematical Foundations of Quantum Mechanics.* Priceton, NJ: Priceton University Press.

[von Neumann 66] von Neumann J. 1966. *The Theory of Self-reproducing Automata.* Urbana, IL.: University of Illinois Press.

[Minsky 85] Minsky, M. 1985. *The Society of Mind.* New York, NY: Simon & Schuster.

[Pask 58] Pask G. 1958. Physical Analogues to the Growth of a Concept, *Mechanization of Thought Processes.* Proc.of a Symposium, National Physical Laboratories, November 1958, London: HMSO.

[Pask 61] Pask G. 1961. *An Approach to Cybernetics.* London: Hutchinson.

[Pattee 92] Pattee H.H. 1992. The Measurement Problem in Physics, Computation, and Brain Theories, in Carvallo M.E.(ed.) *Nature, Cognition and System II.* Dordrecht: Kluwer, pp. 197–192, 1992.

[Pattee 95] Pattee H.H. 1995. Artificial Life Needs a Real Epistemology, in Moran F., et al.(eds.), *Advances in Artif. Life.* Berlin: Springer, pp. 23–38.

[Prem 95a] Prem E. 1995. Grounding and the Entailment Structure in Robots and Artificial Life, in Moran F., et al.(eds.), *Advances in Artificial Life.* Berlin: Springer, pp. 39–51.

[Prem 95b] Prem E. 1995. Dynamic Symbol Grounding, State Construction, and the Problem of Teleology, in Mira J. & Sandoval F.(eds.), *From Natural to Artif. Neural Computation.* Proc. Int. Workshop on Artif. Neural Networks. Berlin: Springer, pp. 619–626.

[Prem 97a] Prem E. 1997. The behavior-based firm. *Applied Artificial Intelligence.* **11** (3), pp. 173–195.

[Prem 97b] Prem E. 1997. The implications of embodiment for cognitive theories. *Austrian Res. Inst. f. Artif. Intell.* Vienna, TR-97-11. `http://www.ai.univie.ac.at/papers/oefai-tr-97-11.ps.Z`

[Rosen 78] Rosen R. 1978. *Fundamentals of Measurement and Representation of Natural Systems.* New York: North-Holland.

[Rosen 85] Rosen, R. 1985. *Anticipatory Systems.* Oxford, UK: Pergamon.

[Rosen 91] Rosen R. 1991. *Life Itself.* New York: Columbia University Press.

[Schrödinger 67] Schroedinger E. 1967. *What is Life and Mind and Matter?* Cambridge, UK: Cambridge University Press.

[Tanji et al. 94] Tanji J., and Shima K. 1994. Role for supplementary motor area cells in planning several movements ahead. *Nature,* 371 (6496), pp. 413–416.

[von Uexküll 28] von Uexküll J. 1928. *Theoretische Biologie.* (Theoretical Biology.) Frankfurt/Main: Suhrkamp.

[Wittgenstein 53] Wittgenstein L. 1953. *Philosophische Untersuchungen.* (Philosophical Investigations.) Frankfurt/Main: Suhrkamp.