Semiosis in Embodied Autonomous Systems

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

Abstract

This paper [Prem 98] discusses processes of semiosis in embodied autonomous systems such as behavior-based robots or animals. The starting point for this investigation are the peculiarities of embodied autonomous systems, i.e. the fact that they are physical systems with a body that is to be moved around in the real world without the help of a human supervisor. We revisit previous results about the nature of representation in such systems. We draw parallels with the philosophical work of Martin Heidegger and show the relevance of these accounts for the study of autonomous sign users. It will be argued that signs are a type of equipment for such systems that reveal a specific interaction context and serve to orient autonomous systems at specific action circuits. These considerations shed new light on a considerable amount of previous work about the usage or "communication" of signs in the field.

1 Introduction

Processes of semiosis are amongst those phenomena which have driven scientific interest with an extreme intensity for a very long period. It is only natural that nowadays this interest in semiotics continues in the field of information systems, where the themes range from iconic visualization, the semiotic processes in the Internet to the symbol grounding problem and robot communication. The purpose of this paper is not to present final results about this difficult problem area, but to fruitfully contribute to this discussion, albeit with a critical tone.

The study of information systems is not only motivated by technical questions. The computer has long turned into a general metaphor for other systems that can be described as processing information, most prominently also the human animal. Obviously, once such a perspective of humans has been gained, it is only natural to concentrate on the information aspect of such systems alone. The claim that many interesting questions about human intelligence can be studied by inquiries into the nature of computers and their programs used to lie at the heart of Artificial Intelligence (AI) and was often argued for [Simon 80, Simon 81].

Discussions of the nature of signs in this context can be mostly attributed to one of two frameworks. The first is to regard the sign as a *means of information transfer* embedded in a process of communication that passes through the following steps: *source-send-channel-message-receive* [Eco 77]. This view is most closely connected to information transfer in the technical sense and to a simplistic view of language. The second, more philosophical position regards signs as bearing an inherently triadic structure in the semiotic process of designation. There are many versions of what exactly this structure may look like, but it always concerns the sign, a referent or sense, and a significant or object. This latter view of signs has only more recently interested scientists in AI in the context of the *symbol grounding problem* and is related to natural language through the "meaning of words" [Harnad 90, Prem 95a, Prem 95].

The symbol grounding problem is originally based on Searle's *Chinese room argument* and consists in the question as to how it is possible for signs (in computer programs) to acquire and possess original meaning, i.e. meaning that is not solely derived from the meaning in the human interpretor's head [Searle 80]. As a reply to Searle and Harnad, many research activities centered around finding solutions to the symbol grounding problem. During a certain period (mainly from 1980-1990) it was believed by many[1] that the problem could be solved by equipping a system with the possibility to learn the meaning of a sign inductively [Harnad 93]. For example, an information system could be presented with images of dogs and the symbols 'dog' and thus learn to connect the two. However, such approaches were soon criticized and dismissed as technically, scientifically and philosophically inadequate [Bickhard & Terveen 95, Prem 95b].

Basically, many of the arguments center around the notion of *representation,* which was originally assumed to lie at the heart of natural and artificial intelligence. The idea was that an adequate representation of the world inside a system would be a necessary prerequisite for its intelligence. The representational items in turn, would be labeled (by symbols) and could easily be expressed as words

[1] including the author of this article

(for descriptive communication purposes) or used in internal models of the world (for planning and purposeful interaction). The prominent work of [Brooks 91] questions the traditional view of representations and symbols and lead to a new interest in embodied autonomous systems.

In this paper, we investigate implications of embodiment and autonomy for semiotic processes. It will be suggested below that the simple view in which internal representations are labeled and spelled out through symbols or names is not a sufficient explanation of semiosis in such systems. These accounts often suffer from a severly restricted view of the purpose with which systems use signs. Additionally, they often neglect the complex underlying representations that are themselves formed by purposeful system-environment interaction.

2 Representation in Embodied Autonomous Systems

2.1 Representation and Action

For the purpose of this paper a system is called *embodied* simply when it has a physical body that is (in part) controlled by the system itself and that dynamically interacts with the world. It must be stressed that the mere fact that a system has a body dramatically changes its system-theoretic properties. For example, a robot cannot be fully described in only information-theoretic terms. It is a physical structure to which the laws of physics apply when it interacts with the world.[2] Robots, for example, are subject to environmental time, while the description of computers needs only talk about internal state-transition time. Such robots are difficult to simulate and thus need to be built in order to be understood and tested. Once a robot is built, the realm of the purely formal is left. The physical interaction of the robot with its world leads to material truth conditions when control systems evaluate sensors [Cariani 90, Pattee 95, Prem 95a].

In the following we call those systems *autonomous* that are able to sustain their purposeful interaction with the world for some time. Obviously, this criterion is not a sharp, clear-cut mathematical definition but allows for *degrees* of autonomy. In the political sense "autonomy" refers to the ability to decide upon one's own affairs but also to the possibility to really do so. Within the robot world autonomy means the very same, namely "to make a living on one's own".

An important parallel between robots and animals in this context is the fact that *both systems are fundamentally alone.* Robots must make decisions without the help of any human supervisor or other being that can be completely relied on. It is the roboticist's task to develop a system design that anticipates the problems, which a robot will have to face. To a large extent, such a solitude is one of epistemic relevance: the system must decide on its own, what there is in the environment and what can be used to pursue its own goals [Prem 97]. An important concept

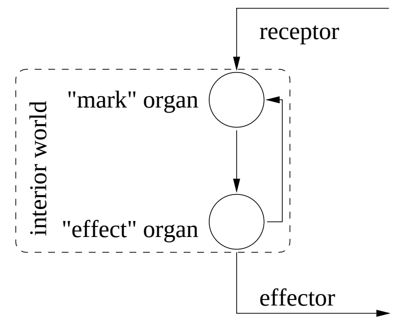

Figure 1: The action circuit, first described by von Uexküll. The properties of the mark organ and of the system-environment interaction determine the interior world of the animal. The dynamics itself, however, is driven by the interaction "purpose", i.e. the internal outcome of the interaction.

in this context is the "action circuit" that was originally introduced by the theoretical biologist Jakob von Uexküll [von Uexküll 28]. Fig. 1 depicts Uexküll's original idea in which the interior world of the animal is equally dependent on its sensory and effector organs, but also on the interactions of the animal with its environment. The purpose of this concept is to better understand the animal's behavior in its natural habitat.

As an example, consider the hungry tick, which is known to bite everything with a superficial temperature of 37° C that emits butyric acid. For the tick there are no humans, deer or grass. All that matters (in the feeding context) are temperature and chemical concentration. In order to understand the tick, we first need to comprehend the properties of the sensory organs. However, it seems equally important to understand the purpose of environmental interaction, since we can easily imagine other action circuits that might exist in parallel for fleeing, mating, etc.

The concept of the action circuit can add much to a notion of representation. [Bickhard 97] pointed out that a useful notion of representation is based on the anticipations of interaction outcomes. For instance, let us assume that there is some part of the nervous system in the tick that is able to relatively reliably indicate the presence of food. It would then be possible to call this indication a *representation* of the tick's world. Several points must be noted here. Firstly, the indication may be one of food in human terms, but the concept formed by the tick is, of course, a completely different one. It means the class of

[2] It is trivial that the same is true for regular computers. However, they are built in a way that the physics that drives the programs is highly constrained and controlled by a well-known set of rules.

all things with a given temperature and chemical diffusion. Secondly, and more importantly, this representation is based on *interaction outcome.* We may assume that it was formed in a long process of evolutionary adaptation or learning. It thus bears a character of anticipation in it. The anticipation consists in the prediction of a specific interaction outcome, *internal* to the system that is coupled to the representation. The activation of a representation of "food" indicates that the specific interaction of "biting" now may appease hunger. Most notably, such a representation may as well be wrong. This is the case, when the anticipated interaction outcome does not occur.

Note that the kind of actual interaction with the world plays a major role in the kind of representation that becomes active. We may well imagine a situation (e.g. mating) in which the indications of biting outcome may not be important at all. In these situations the "food" remains opaque to the tick. It is straightforward to imagine every object that an animal may be able to distinguish as the outcome of such a process of adaptation. In other terms, "things" in the environment of an autonomous system can be regarded as formed by anticipations of interaction outcomes or as "tools" to support the purposes of environmental interaction of the autonomous system.

2.2 Representation and Ontology

While animals and especially robots may have rather restricted ways of interactions with the objects around them, humans are much more versatile. For example, we may use a pen as a pointer, for picking, or simply in order to write. How the thing will look to us, will primarily depend on our current engagement, i.e. on the context of our current activities. The purely physical properties of things such as optical, haptic, or structural features will only rarely be of primary importance to us. Usually, we just take the pen and start writing or pointing in a most natural way of being engaged in interactions with the world around us. The encountered entities, such as pens, seem to become different things depending on the context in which we encounter them.

It should be emphasized here that this ontology of things as tools in interaction circuits is very similar to the philosophy of Martin Heidegger. [Heidegger 27] argues that a tool-like character lies at the very heart of everyday coping with the world. We usually do not need to deliberately think about how to use the objects around us. It is much more the other way round: When we are engaged in a certain interaction, things "show up" as solutions to our problems. I suggest that they indicate the corresponding interaction outcome because they were formed as anticipations of these internal interaction results.

For Heidegger, these elements of equipment around us also bear a network of references to related interactions and

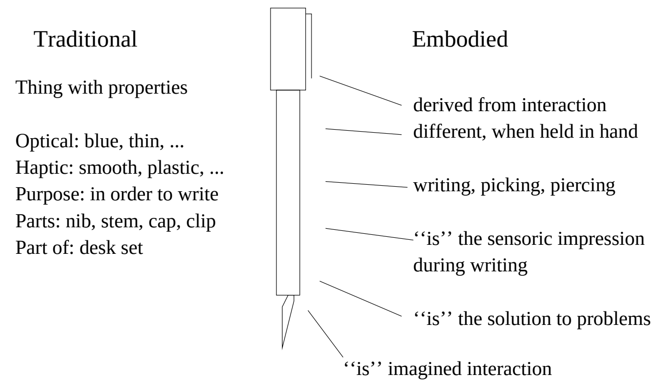

Figure 2: Differences between a pen in the traditional and in an embodied perspective. In the latter, the being of the pen is depend on the context in which it is encountered. It is derived from purposeful interaction with the world.

other equipment. For example, the pen refers to the paper and to writing. The well-known AI critique of [Dreyfus 72] is based on the observation that the old approach to AI models of the mind using symbolic representations becomes awkward once such a different ontological view is accepted. The traditional description of this kind of tools would be in terms of physical and other properties, not starting from its functions but trying to derive them from these properties (see Fig. 2, cf. [Dreyfus 90]). In our view here, the functions lie at the very heart of *what things are,* they are the basis for the kind of entities that autonomous systems encounter in their world.

The important point is that this analysis holds true for every system of a certain degree of autonomy that is capable of dynamic interaction with the world. As soon as such a system adapts to its environment (by learning or by an evolutionary process), this results in the kind of anticipated interaction outcomes mentioned before. [Rosen 85] shows how such an adaptation generates a model of the selection (or adaptation) process within the system and thus results in internal "representations" (in the technical sense) that indicate future successful interactions (cf. [Prem 97]).

3 Signs as Tools

3.1 Processes of Semiosis in Autonomous Embodied Systems

Let us now assume that the embodied autonomous system under investigation is no longer alone, but shares the world with other systems. From the viewpoint of the above ontological understanding of embodied autonomous systems, we gain a new understanding of sign use in this context. In the same way that a pen may be regarded as a tool for writing, signs are tools for indication purposes. The famous example in a human context is the turning signal of a car [Heidegger 27, p.76f]. The signal is used by the driver in order to indicate the driver's intended direction and to carefully inform others about this planned action. Other drivers take up this signal and hopefully behave according to it. In fact, we may well say that the sign is primarily used by the other participants in the traffic.

Using signs in this way is not restricted to human beings or higher animals. From the viewpoint of embodied autonomous systems it can also be found in simple insects, e.g. in forms of stigmergic communication of ants that indicate places and directions to other members of the species. Signs of this kind can also be used for oneself, e.g. to find the place in spring where the nuts were hidden before winter.

The appropriate reaction to a sign encountered in the environment is to "give way" or to "direct oneself to" or to "start digging at". It would usually not be an appropriate behavior for an animal to simply stare at the sign trying to "grasp its meaning". As Heidegger rightly points out, the more the sign is actually used, the less it becomes visible as a mere sign. Instead, it serves to orient the system's interaction, for example at a different object.

Note that this view is very different from the traditional semiotic approaches, e.g. the semiotic triangle. Primarily, we are not at all interested in what a sign may *denote* or what its "meaning" may be. Instead, we study the *orientation process* that happens when signs are encountered in the world. In this view, signs of this type are a kind of equipment for autonomous systems that reveal a specific interaction context. Signs show the "wherein of living." Signs orient action circuits.

As a well-known example, consider the dances of honey bees which indicate distance and direction of food sources to the whole population. The direction which is encoded in the orientation of the dance path is directly taken up by other bees. Their action circuit "search for food" becomes oriented at the "intended" direction. Indeed, the whole control process seems to be similar to a program-driven system. The communicated action parameterizes a relatively fixed action pattern.

The interesting aspect of these processes is that they make the aforementioned web of references of equipment explicit. The turning sign not just points to a specific direction and "represents" it in a symbolic way. The interactional whole of driving, taking care, turning etc. gets its new orientation from the usage of signs provided by others.

This implies that processes of semiosis are rather different in nature from previous accounts. Signs are not mere entities in relation to other entities to which they stand in a specific "reference-to" relationship. Instead, in autonomous systems research, signs should be regarded primarily as tools for sign users in environmental interaction for themselves or for others. Again, these signs are tools formed by successful interaction that consists in "showing" or "pointing to" and results in the desired outcome for the sign users (which now includes two agents). Bringing this argument to the extreme means that *signs are anticipations of successful interactions of indication.*

3.2 Autonomous Sign Using Robots

Recently, there is an increasing interest in the study of robot communication. This line of research was often motivated by an interest in the origin of language or in (child) language acquisition. Many researchers start with the triadic view of symbolic reference. For this, ways to negotiate the form of signs between agents and to discover some common "meaning" of these signs must be developed. This view is tightly connected to the symbol grounding problem and to the communication view of sign use. In the latter the problem consists in protocols to transfer a given or intended "meaning" to another agent [Prem 95c]. Often the underlying methodology consists in a two-step approach to the problem. First, the common "meaning", i.e. the concept itself is developed or grounded. In a second step, agents construct a system of linguistic reference to their internal, often "private" concepts. This is the "language learning" step, in which systems give names to concepts. More recently, it was widely acknowledged that these two processes may actually be interdependent and influence themselves in technical approaches with robots as well as in language learning experiments with children.

Our approach suggests that a large part of this previous work must be criticized with respect to the underlying assumptions. Although we have not said much about linguistic signs here, many oversimplifications in experiments about the "origin of language" or the "foundations of meaning" are revealed as soon as processes of semiosis are regarded as centered around purposeful environmental interactions. Such a viewpoint immediately emphasizes the dynamic aspect of sign use as driven by interaction outcome. This further entails a move away from some "outer" entity that becomes mapped in internal representation. A correct treatment of robot herd communication in this context would have to start from regarding communication as motivated by the autonomous system's social interaction purposes. Accordingly, the first word "spoken" by such a system might well be "go away!" or "give me your food!" or "I love you!" and not a label for "food source" or "enemy".[3]

The study of these origins must also describe processes that underlie sign use, namely the foundational process of taking something as a sign in order to get orientation. Heidegger's example here is the wind from the south that the farmer discovers to be a sign (Peirce: "index") for rain.

[3] In practice, however, it may well be difficult to distinguish between the label "enemy" and the warning call "enemy!" as an anticipation of a hopefully successful escape action.

It just is not the case that some existing physical entity is "taken for something else". Instead, the farmer's purposeful interaction with its environment *discovers* the wind from the south in this way of being an indication of rain. This example shows that–primordially–a sign cannot have just about any meaning but only that which "makes sense". Symbolic tokens may well be arbitrary in a certain sense, however their meaning will be tied to purposeful interaction and to the social aspects of sign-based interaction.

Note that all these processes of semiosis remain at the level of mere signs, *not* of language so far. It is only natural to first develop a clear theory of (i) how autonomous systems create signs and (ii) socially use signs in order to produce system-internal outcomes. Only in later steps should we try to describe a complex system of communicative skills that can serve many different purposes from telling jokes to asking for help.

3.3 More Philosophical Investigations

A critical point of our approach is that these preliminary signs may look just like about anything else encountered in the environment of an autonomous system. Consider the example of a robotic system that learns to detect signs in order to decide whether to turn left or right for food (e.g. [Ulbricht 96]). Unless such a system starts a discussion about "the meaning of signs" it will just seem to use them in an everyday fashion of coping with the world. The search for a criterion that allows us to distinguish sign users from other agents interacting with the world looks very much like a Wittgensteinian "investigation" ([Wittgenstein 53]): When would it be possible for us to say that the system "really understood" the sign, i.e. which criterion would ensure that it "learned the rule"?

Firstly, it must be kept in mind that signs are used for indications by agents. It follows then that they are used between agents, made up by agents for other agents (or for oneself). This is a major difference to other entities in the environment that may possess the above mentioned tool-character, however are not addressed to other beings. It is this aspect of "addressing" that implies that different interaction contexts must "show up" in sign use. While in our everyday unproblematic usage of tools they remain tacit (i.e. are not revealed to be "tools"-we just use them), this is not true for signs. Signs are supposed to speak to us, they must not remain simple objects, but need to indicate something that they themselves are not.

Secondly, calls for a method to "verify" the correct understanding of signs will force the definition of some purely outer criterion. The reason for this is twofold.

(i) Behavior can be observed and checked. Anything internal to the system ultimately remains concealed. Note that even in the case of robots, where we could trace the representations formed, such a tracking would be extremely difficult due to the dynamic interaction with the world. More importantly, the conceptual framework within the system will be conceptually opaque to the observer since it is the result of a complex process of adaptation, the purposes (or selection criteria) of which we need not even know. Wittgenstein quite clearly describes this fact when stating: "Even if a lion could speak, we would not understand it."

(ii) Signs in our examples here are not words, they are tools for indications. But rather than emphasizing the *indication* element of signs, the sign user will care about the effect that they have on *behavior.* The sign user's primary interest is the generated (re-)action. At least one such reaction will be the *socially* desired one that can be interpreted as "correct" understanding.

4 Conclusion

In this paper we have presented a rather unorthodox view of signs that is motivated by the study of autonomous embodied systems. Such systems are not only information systems, since they are also heavily influenced by physical processes. All the more, this makes them interesting test cases for the study of semiotic processes. We suggest to look into these processes from an existential-ontological viewpoint and regard signs as tools to control aspects of the interaction with other agents.

In *Being and Time*[4] Heidegger lists three properties of the relation between semiotic reference and the tool-based reference that we find in "equipment":

- Reference is based on the equipmental "in-order-to", i.e. on the tool-character of signs.
- The indication of signs is part of a referential whole in which signs are embedded like any other equipment.
- A sign explicitly lights up the system of reference in which it is embedded through appropriate usage.

These properties must be understood as a first approximation to the phenomenal character of semiotic referral. They should not be mistaken for "properties" of signs. As Heidegger says, equipment does not possess properties in the traditional ontological sense, it can only be adequate or inadequate for certain types of interaction. The creation of systems which can successfully use signs in this sense would be a fruitful starting point for further inquiries into the nature of semiotic processes in embodied autonomous systems. Dreyfus turns it this way: "Signs can do their job only because we already know our way about in the world." [Dreyfus 90, p.102] or as Heidegger says: "Signs always indicate primarily "wherein" one lives, where one's concern

[4] Sec. 17, p. 82.

dwells, what sort of involvement there is with something." [Heidegger 27, p.80]

On this low level of complexity that we study here, we now find a pronounced *pragmatism*. Although we suggested to leave behind the semiotic triangle and the communication view of signs, it now shines through, how these constructs may come in again. The *meaningful content* of the signs, if this concept is to be kept, lies in the concrete orientation that they give to action circuits of those who already share a world. The "meaning" of the signs they use, i.e. the orientation which they give in practical contexts, is accessible to and can be the subject of behavioral experiments. This however, is something which C.S. Peirce suggested as a strategy to study the meaning of signs long ago [Peirce 31, CP 5.411].

Acknowledgments

The author gratefully acknowledges support from the Austrian Research Institute for AI which is sponsored by the Austrian Federal Ministry for Science and Transport.

References

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

[Bickhard 97] Bickhard M.H.: The emergence of representation in autonomous agents, in Prem E. (ed.) *Epistemological Issues of Embodied AI.* Cybernetics & Systems, 28(6), 1997.

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

[Cariani 90] Cariani P.: Emergence and Artificial Life, in Langton C.G., et al.(eds.), Artificial Life II: Proceedings of the Second Artificial Life Workshop, Addison-Wesley, Reading, MA, 1990.

[Dreyfus 72] Dreyfus H.L.: What Computers Can't Do. Harper & Row, New York, 1972.

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

[Eco 77] Eco U.: Zeichen (Signs), Suhrkamp, Frankfurt/Main, 1977.

[Harnad 90] Harnad S.: The symbol grounding problem. Physica D, 42, 335–346, 1990.

[Harnad 93] Harnad S.: Symbol grounding is an empirical problem. In Proc. of the 15th Annual Conference of the Cognitive Science Society, Boulder, CO, June, p. 169–174, 1993.

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

[Peirce 31] Peirce C.S.: Collected Papers, C. Hartshorne, P. Weiss (eds.), Harvard University Press, 1931.

[Pattee 95] Pattee H.H.: Artificial life needs a real epistemology. In Moran F., et al.(eds.), *Advances in Artificial Life.* Springer, Berlin, 23–38, 1995.

[Prem 95] Prem E.: Symbol grounding and transcendental logic. In Niklasson L. & Boden M.(eds.), Current Trends in Connectionism, Lawrence Erlbaum, Hillsdale, NJ, p. 271–282, 1995.

[Prem 95a] Prem E.: Grounding and the entailment structure in robots and artificial life, in Moran F., et al.(eds.), *Advances in Artificial Life.* Springer, Berlin, 39–51, 1995.

[Prem 95b] Prem E.: Dynamic symbol grounding, state construction, and the problem of teleology. In Mira J. et al. (eds.), *From Natural to Artificial Neural Computation.* Springer, Berlin, 619–626, 1995.

[Prem 95c] Prem E.: Understanding complex systems: What can the speaking lion tell us? In Steels L.(ed.), The Biology and Technology of Autonomous Agents, Springer, Berlin, 1995.

[Prem 97] Prem E.: Epistemic autonomy in models of living systems. In Proc. of the Fourth European Conference on Artificial Life, Brighton, MIT Press/Bradford Books, 1997.

[Prem 98] Prem E.: Semiosis in embodied autonomous systems. Proc. of the ISIC/CIRA/ISAS98, Omni-Press, Madison, WI, 1998.

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

[Searle 80] Searle J.R.: Minds, Brains and Programs, Behavioral and Brain Sciences, 3, 417–457, 1980.

[Simon 80] Simon H.A.: Cognitive Science: The Newest Science of the Artificial, Cognitive Science, 4(1), 33–46, 1980.

[Simon 81] Simon H.A.: The Sciences of the Artificial, MIT Press, Cambridge, MA, 1981.

[Ulbricht 96] Ulbricht C.: Handling time-warped sequences with neural networks. In Maes P., et al.(eds.), From Animals to Animats 4, MIT Press, Cambridge, MA, 180–192, 1996.

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

[Wittgenstein 53] Wittgenstein L.: Philosophische Untersuchungen (Philosophical Investigations), Suhrkamp, Frankfurt/Main, 1986, 1953.