The Value of Meaning for Autonomous Robots

Christopher J. Stanton
Innovation and Enterprise Research Laboratory.
University of Technology, Sydney
cstanton@it.uts.edu.au

Abstract

This paper examines the related problems of meaning and symbol grounding with respect to epigenetic robotics. While symbol grounding aims to give artificial systems "intrinsic meaning", most existing approaches to symbol grounding address meaning as a problem of categorical perception, i.e. a theory of reference for maintaining correspondence between internal representations and external entities at a linguistic level. We argue that reference is only one aspect of meaning, and that nonlinguistic and pre-verbal creatures are also meaning users. As such, we argue meaning plays an important role in an agent's value system, providing intrinsic motivation and reinforcement for life-long development and learning. Lastly, we explore how models of meaning can help shift the intellectual burden of grounding from the programmer to the program by designing robots capable of grounding themselves.

1. Introduction

The founding aim of symbol grounding (Harnad, 1990) is to "ground" symbols so that they can be "intrinsically meaningful"; that is, to create meaning in such a way that the symbols of a symbol system can be meaningful to the symbol system, and not just meaningful to the system designer. To this end, approaches to symbol grounding have attempted to ground meaning through categorical perception, by creating and maintaining internal representations which refer or correspond to real-world entities. However, there is no proof that this strategy will produce intrinsic meaning. This paper argues there is more to a theory of meaning than reference, and that the way forward is to examine the evolution of the brain in order to understand the evolutionary value of meaning. It is argued modeling intrinsic meaning requires modeling intrinsic "value systems", which can enable an autonomous robot to meaningfully evaluate novel experiences.

The structure of this paper is as follows: section 2 considers the related concepts of meaning and minds, briefly reviewing background literature such as the Chinese Room (Searle, 1980) and symbol grounding (Harnad, 1990), and discusses the difficulty in proving the existence of either. Section 3 considers the premise of symbol grounding, that the symbols of symbol system can be grounded in non-symbols, and considers whether "non-symbols" need to be grounded also. Section 4 considers theories of meaning in relation to symbol grounding, noting that most approaches to symbol grounding involve creating and maintaining reference (or correspondence) between an internal model of the world and the world itself. Section 5 considers how meaning relates to agent's value systems. The paper concludes by considering how to model value systems.

2. Background: Meaning and Minds

"Meaning" is important to human beings. We seek "meaningful lives", "meaningful experiences", while some things in life mean more to us than others. Sometimes we even find (supposedly) "meaningless" things to do, just to pass time and alleviate boredom. However, despite the obvious importance of meaning in our lives, understanding what "meaning" exactly is has proved elusive for philosophers and researchers throughough the history of science - we can't even agree whether meaning is a "thing" in our heads, or something out there in the world.

As a human being, it is clear to myself that my thoughts are meaningful to me. It is also clear to myself that I am a conscious being, with intention and emotion. Moreover, I also think it is reasonable to assume that other people's thoughts are meaningful to them, despite having no method or means to actually prove it. Likewise, I suspect my dog actually "thinks"[1], and that some things in her life are more meaningful than others (for example, she definitely appears to be "happy" and "excited" when I get home from work) - but again, unfortunately I have no way to prove it.

Now, this brings me to my robots. Sometimes

[1]However, I suspect her thoughts and mental processes are quite different to mine.

(with the "right" or "appropriate" programming) - they behave in ways that appear (crudely) intelligent, such as performing tasks, overcoming minor problems, even learning small things along the way. The Sony AIBOs in our lab can even play a primitive version of "fetch" like a real dog, albeit on a much smaller, poorer and slower scale.

Do these robots have a "mind"? Do they possess "meaningful thoughts"? As I programmed these robots, I am convinced they have no meaningful thoughts, emotions, or intentions - they are mere computers connected to sensors and motors, blindly executing my programmatic instructions. Whereas for other people and other animals I am unable to prove the existence of their minds, the converse is indeed true for my robots - that is, I can not prove these robots do not have a "mind", regardless of how simple or well understood their inner workings are.

2.1 Trapped in the Chinese Room

Where does this leave us? The Chinese Room (Searle, 1980) famously argued that a sophisticated program, capable of passing a Chinese version of the Turing Test, would have no understanding of Chinese as it is simply following the instructions contained in a rule-book. To make his point Searle imagined he was the computer. Trapped inside a room, notes with Chinese characters are passed in through a small hole to Searle. Searle then diligently uses a rule-book to find the input characters (based upon their shape), and then scrawls down the corresponding output set of Chinese characters and passes the notes back out. To the Chinese observer on the outside of the room, it appears the man inside the room can understand Chinese, but Searle argued the opposite is true - he understands nothing of Chinese (the characters are "meaningless shapes"), as he is just following the rule-book.

Searle's simple thought experiment has provoked enormous debate. Can a computer have a mind? Can a computer "understand"? What are "meanings"? Where do meanings exist? Searle's Chinese Room is an argument against "strong AI" - proponents of which hold that cognition is computation, i.e. a mind is the result of the right "program". Searle argues that cognition must involve more than simply computation (i.e. semantics). As Searle has no understanding of the notes he is receiving and passing out, Searle's view of meaning is that of semantic internalism - that meanings exist "in the head" or mind of the meaning user. In contrast, semantic externalism holds that meanings exist in the world, e.g. see (Putnam, 1973)[2].

I can use the Chinese Room argument to support - but not prove - my belief that my ball-playing AIBOs have no mind or meaningful thoughts. However, a common reply to the Chinese Room argument is the long-standing "other minds" problem (Harnad, 1991) - how can I prove that any other living entity has a mind? Other people may behave as if they have a mind, but this is not proof of a mind[3]. So, if I can't prove other people or animals have minds, how can I prove a robot has or doesn't have a mind? Furthermore, if we can't prove the existence of human minds, how can we prove or deny the existence of artificial minds? There is no forseeable solution to this dilemma.

2.2 Is escape possible? (via symbol grounding)

Simply put - despite our differing personal intuitions, hunches and suspicions - we are unable to prove or disprove the existence of either natural or artificial minds, yet countless keystrokes have been spent debating whether a computer can have "consciousness", "understanding" or a "mind". Are we wasting our time? Are we missing the point?

"Symbol grounding" (Harnad, 1990) has offered hope of escape to many. Motivated by the Chinese Room argument, Harnad likened the symbol grounding problem to trying to learn Chinese from a Chinese dictionary alone, where every word is defined in terms of other Chinese words. Thus, without knowledge or experience of any Chinese words, as each Chinese word is defined in other meaningless Chinese words, the task is seemingly impossible. To avoid this infinite regress, Harnad suggested that the symbols of a formal symbol system "must be grounded bottom-up in 'nonsymbolic' representations of two kinds", namely iconic and categorical representations. Harnad described iconic representations as "sensory projections" which structurally resemble the thing they represent (e.g. the iconic representation of a horse would be the shape a horse casts on our retinas), while categorical perception learns the "invariant features" of sensory projections.

Harnad touted the emerging field of connectionism as an ideal candidate for learning the "nonsymbolic" representations, thus advocating a hybrid system, in which the nonsymbolic iconic and categorical representations would provide an escape from the infinite regress of the Chinese/Chinese dictionary (i.e. the formal symbol system). Harnad argued this approach would provide "intrinsic" meaning to the

[2]For example, the "Twin Earth" thought experiment (Putnam, 1973) asks us to consider a mirrored, twin earth somewhere in the universe where everything is exactly the same, except the chemical composition of water. On each world there are two beings, both who possess a concept about "water", but because the chemical composition of water is different on the two earths - and thus the referents of their respective representations are different - Putnam argued the meanings of "water" are different, and thus meanings "ain't in the head".

[3]i.e. they could be "philosophical zombies".

symbol system, by providing an internal "semantic interpretation" (or grounding) of the symbols of a formal symbol system. However, Harnad also hedged his bets, conceding in conclusion that there is "no guarantee that our model has captured subjective meaning".

A large body of multi-disciplinary research has been generated by the grounding problem. While grounding-related research varies in implementation details and application area, most approaches have focused on ascribing meaning through categorical perception of sensorimotor experience. To this end, numerous types of hybrid systems have been developed in which different techniques are used for classification and categorisation of sensory input data. For example, logic (for high level reasoning) has been complemented by different methods of categorisation and classification, such as different neural network based approaches (Harnad, 1990, Davidsson, 1993, Nakisa and Plunkett, 1998, Riga et al., 2004, Law and Mikkaulainen, 1994), genetic and evolutionary algorithms (Swarup et al., 2006), self-organising maps (Blank et al., 2002), and the geometric based representations of conceptual spaces (Gärdenfors, 2000, Chella et al., 2003).

An underlying analogy and assumption behind these approaches is that the output data produced by a robot's sensors is somehow comparable to the sensations we feel (such as taste, touch, etc) - a form of "robot functionalism" (Harnad, 1995). In other words, I know what a pizza means because I know what it tastes like, smells like, looks like, feels like, and so forth - and therefore, if we could do the same for a robotic system (using cameras, taste sensors, etc), the robotic system's meaning of pizza would also be grounded (Mayo, 2003).

Despite the large body of grounding related research generated over the last 20 years, the question of whether any of these systems capture "intrinsic meaning" or "subjective meaning" is still unanswerable. Has any computational system ever escaped the Chinese Room? Do any of these systems exhibit understanding? Approaches to symbol grounding have (of course) failed to prove they capture "subjective" or "intrinsic" meaning - the Chinese Room argument still applies, regardless of whether logical formalisms are connected to other forms of "nonsymbolic" representation. Computers are still blindly following their underlying programmatic instructions, regardless of the higher level formalism (i.e. the implementation's conceptual details).

3. Grounding symbols with nonsymbols?

> "The meaning of a representation can be nothing but a representation. In fact, it is nothing but the representation itself conceived as stripped of irrelevant clothing. But this clothing never can be completely stripped off: it is only changed for something more diaphanous. So there is an infinite regression here" (Peirce, 1935).

Is it possible to give meaning to a "symbol system" through a "nonsymbol system"? This is the founding premise of symbol grounding. It raises many questions. What is a symbol, and what is a symbol system, and do other types of representation also need grounding? What is meaning and why is it important? Does "grounding" really give or create "meaning"? A program is just a program, regardless of the higher-level formalism or conceptualisation.

In defining the symbol grounding problem, Harnad describes the problem in relation to "symbolic AI". However, what is a symbol? A symbol designates or denotes something else (Newell and Simon, 1976), while a "representation", by definition (i.e. to *re-present*), stands for something else (the referent). Is there a difference between a symbol and other non-symbolic types of representation? Harnad offers a detailed definition of a symbol system. According to Harnad, the main difference is the arbitrary nature of symbols and the non-arbitrary (or causal) nature of non-symbols. For example, a symbol can be likened to an arbitrarily assigned label, whereas the nonsymbols that (supposedly) ground symbols are causally related to things they represent. For example, iconic representations are structurally related to the entity they represent, e.g. the shape of a horse cast on the eye's retina.

Steels (2007) comments the use of the term "symbol" in artificial intelligence has probably created "the greatest terminological confusion in the history of science". Steels argues this confusion arises from differing uses (or meanings) of the term symbol by researchers with different backgrounds, i.e. philosophers, linguists and computer scientists use "symbol" in different ways. For example, a computer programming language is itself symbolic, yet when a neural network is implemented in a computer language using such symbols, the neural network is not considered by cognitive scientists or philosophers to be "symbolic" - rather, it is considered to be "subsymbolic". Many years ago Chalmers (1992) made a similar point:

> "In talking about symbol grounding, one must be careful here not to fall into the old trap of conflating the two quite separate meanings of 'symbol', that is, 'computational token' and 'representation'. When one argues that computation consists in the manipulation of meaningless symbols, this is a point about computational tokens. When one asks

how symbols can be grounded, this is essentially a question about representations." (Chalmers, 1992)

Chalmers (1992) is not alone. Several other authors (MacLennan, 1993, Pfeifer and Verschure, 1995, Ziemke, 1999), have commented that the symbol grounding problem is not limited to "symbols", but to representation in general. For example, MacLennan (MacLennan, 1993) describes the grounding problem as "how do representations come to represent", while Pfeifer and Verschure (Pfeifer and Verschure, 1995) describe a "general grounding problem" which applies to knowledge "structures", rather than just "symbols". Lastly, "theory grounding" describes how beliefs about the world need to be grounded through embodied interaction with the world (Prince, 2001).

The Chinese room argument raises the issue of how any computational machine can understand or be a meaning user, as a computational machine will always follow a set of instructions (i.e. the synactical rules of the program). If the grounding problem applies to all computational tokens, how can grounding symbols in non-symbols make symbols intrinsically meaningful to the program? Symbols in symbol grounding research usually refer logical propositions or linguistic expressions that are used for reasoning about the world. At one extreme, there are some researchers who believe grounding only applies in the context of communication, e.g. "a symbol is a convention between two (or more) agents. Thus it makes no sense for a single agent to try to ground symbols" (Swarup et al., 2006). However, the vast majority of animals are pre-verbal creatures who communicate very little. Do they have symbols? There is evidence that many animals have representations (Gärdenfors, 2003). Are their representations intrinsically meaningful? At the least, they are useful. Moreover, there are many other ways of reasoning than by using a formal (logic-based) symbol system. In robotics (and software systems in general) many different types of representation are used for decision making - databases, classes, variables and so forth are all used to refer to things in the real world. Remember - the premise of symbol grounding is that meaning can be attributed to a symbol system through nonsymbols. But if all computational tokens need grounding, where do we do we get off the grounding merry-go-round?

4. Grounding and Meaning

4.1 Intrinsic Meaning? Or a Theory of Reference?

"Symbol grounding is a new name for an older problem - the problem of providing a theory of reference for atomic formulae of a system of internal representation"(Christiansen and Chater, 1993)

Harnad's appeal to the Chinese Room and the notion of intrinsic semantics concern philosophical issues such as intentionality, consciousness and meaning. Harnad's practical solutions (connecting symbols with sensorimotor experience of the real-world entity referred to by the symbol), however, attack a very different problem - a theory of reference (Christiansen and Chater, 1993). A theory of reference (or alternatively "correlational semantics" (Prem, 1995)) involves relating internal representations with external entities (e.g. somehow connecting a symbol "John" with the real-world "John") - in other words, understanding how the atomic units of a language come to have meaning.

A large proportion of grounding-related research treats meaning as a problem of reference. For example, symbol grounding has been described as establishing the "direct correspondence between internal symbolic data and external real world entities" (Albus and Barbera, 2005); the problem of how "symbols should acquire their meaning from reality" (Vogt, 2002), or the association of a symbol "with a pattern of sensory data that is perceived when the entity that the symbol denotes is seen, or tasted etc" (Mayo, 2003). "Anchoring" (Coradeschi and Saffiotti, 2000), a variation of the grounding problem, embraces the problem of reference for physical objects - anchoring involves "maintaining the correspondence between symbols and sensor data that refer to the same physical objects" (Coradeschi and Saffiotti, 2003).

Much of the emphasis on reference-based meaning in grounding literature may rest with the important role it plays in language, as for communication to occur it is imperative the communication participants have shared meaning, i.e. that they establish "common ground" (Clark and Schaefer, 1989). Indeed, there is even a branch of grounding related research called "social symbol grounding", a term used to describe the formation of shared meanings through communication (Cangelosi, 2006) - the process of multiple agents using the same term to refer to same real world entity. There have been many examples of such language-grounding work within robotics which examine how language may evolve. For example, the "Talking Heads" experiments (Steels and Vogt, 1997, Steels and Kaplan, 2002) involved two cameras interacting in a simplified visual environment which consisted of coloured shapes on a white board. In these experiments, the agents develop a shared lexicon for the entities in the environment (i.e. the coloured shapes) through the use of language games.

Undoubtedly, maintaining a faithful reference between an agent's beliefs about the world and the reality of that world is important. Systems from airline reservation databases to autonomous mobile robots rely on well grounded representations. For example, an airline reservation system must manage information about flights and passengers in a way that corresponds to real flights and real passengers. Similarly, an autonomous mobile robot that navigates a physical space will be more effective in achieving its objectives if its internal representations of physical barriers correspond to real physical barriers in its environment. Moreover, different agents "groundedness" (Williams et al., 2005) will vary, i.e. a robot's model of the world will invariably have a degree of error, and the nature of that error can vary. For example, a robot soccer player may miscalculate the distance of the soccer ball from the robot, or alternatively fail to "see" the ball entirely.

4.2 More than reference

Ensuring symbols and representations about the world refer faithfully to that world is an important aspect of grounding. However, referents aren't the sole source of meaning for symbols. For example, Harnard argues:

> "We know since Frege that the thing that a word refers to (its referent) is not the same as its meaning. This is most clearly illustrated using the proper names of concrete individuals (but it is also true of names of kinds of things and of abstract properties): (1) "Tony Blair," (2) "the UK's current prime minister," and (3) "Cheri Blair's husband" all have the same referent, but not the same meaning."(Harnad, 2003)

If we consider linguistic meanings of meaning, referential meaning (e.g. "I *meant* that one!") is just one component. Other aspects are:

- *Understanding* (e.g., as in "do you know what I mean?"). Thus, meaning is often defined circularly as how an event, action, word, etc is *understood*, and that conversely, to understand something is to know the *meaning* of it. For example, Barsalou(Barsalou et al., 1993) describes meanings as "people's understandings of words and other linguistic expressions". Likewise, it is this "understanding-based" notion of meaning that is appealed to by Searle(Searle, 1980) in his Chinese room argument - the man inside the room does not understand the Chinese symbols, and thus they are meaningless to him;
- To imply *consequence* or causation (e.g. "that alarm means trouble", "friction means heat", or "lower costs mean lower prices");
- *Intention* - we use the term "meaning" with regard to discussing intent, design or purpose (e.g. "I meant 8am - not 8pm!", "I meant to go swimming this morning, but I overslept", or "that building is meant for storage").
- And lastly, *value* - the worth, relevance or significance of something to the something else (e.g. "the critic's opinions meant nothing to the author" or "her boyfriend meant a lot to her");

It is this last meaning of meaning - the ability to qualitatively evaluate the worth or value of things and experiences - that has been largely ignored in symbol grounding research. In the following section the concept of "meaning as value" is explored in further detail.

5. The Value of Meaning

If we assume the brain evolved, and our experience of meaning is a cognitive function, then it follows that meaning serves to guide an organism's survival. While most grounding research has focused on high-level symbols (such as language), pre-verbal beings such as animals and children interact meaningfully with their environment. This meaningful interaction is the product of internal value systems - mechanisms for judging and discriminating "good" experiences versus "bad" experiences. For example, having a full stomatch is preffered over an empty one; a warm bed is preferred to a cold one; a state of safety is preferred to the presence of a predator (Cisek, 1999). These innate preferences and values provide motivation for behaviour. We search and strive for desirable states, and actively avoid unfavourable ones. Thus, Cisek (Cisek, 1999) argues computers make "poor metaphors for brains", as there is "no notion of desirable input within the computing system", which leads to the "riddle of meaning". Zlatev (Zlatev, 2001) describes meaning as a "a relationship between the individual and the environment, picking out the categories in the environment which are of value for the individual". Likewise, Ziemke and Sharkey (Ziemke and Sharkey, 2001) compare artificial systems with living systems, noting one of the reasons today's artificial machines lack intrinsic meaning is due to their lack of an inntrinsic life task.

It has been argued that for any computer or robot to display intelligence it will require the ability to evaluate experience in a qualitative manner, and perhaps even emotion (Sloman and Croucher, 1981). While emotions may be unneccessary, autonomous robots will require a value system. Value systems are often described in the context of developmental and epigenetic robotics, in which a value system is used as a source of reinforcement for learning and motivation, signalling the occurrence of salient sensory inputs and for qualitatively evaluating the effects of actions.

5.1 Intrinsic Meaning and the Chinese Room

The Chinese Room argument can be twisted to demonstrate the point made in the previous section - that meaning is linked to an agent's value system, which exists to guide exploration and interaction with the world. In Searle's thought experiment, everything that happens to Searle is meaningless and inconsequential. The Chinese symbols are meaningless to Searle, not just because of his lack of knowledge of Chinese, but as a result of the lack of any other salient, significant or relevant events. In reality, why would anyone even bother to follow the rule-book? As Searle highlights, human beings are intentional creatures, and would not simply follow a rule-book unless there was some incentive for doing so. Human beings are driven by internal, intrinsic motivators, and rarely (if ever) act without reason.

Now, consider if at some point in time, as Searle keeps methodically producing his Chinese characters from his rule-book, some delicious food was suddenly passed into the room. Immediately, the recent characters output by Searle would have some meaning to him. If Searle was hungry, it is likely he would try repeating the previous output characters to discover if more food arrives in his prison. If this action produced more food, Searle's hypothesis that the particular pattern of characters means "give me food" would be strengthened. If it didn't, the hypothesis would be weakened, and Searle would probably consider other possibilities as to what caused the food to appear inside the Chinese Room.

Currently, robots are like Searle in the Chinese Room. Everything that happens to them is *meaningless*. There is no intrinsic value in their experience. By endowing robots with (a model of) an intrinsic value system, novel experiences and interactions with *any* environment can be evaluated *meaningfully*. For example, consider current approaches to machine learning - each learning task is small in scope. The value judgements are provided by human supervisors in most cases, and if not, the mechanism for providing unsupervised reinforcement is carefuly crafted.

6. Value Systems

Value systems in biological systems mediate environmental saliency and modulate learning in a self-supervised and self-organized manner. In the mammalian brain, the output of the neuromodulatory system acts as a value signal, modulating widely distributed synaptic changes. Neuromodulators are chemical transmitters in the brain that can have a strong and lasting effect on behaviour. The neuromodulatory systems include noradrenergic, serotonergic, dopaminergic, and cholinergic projections from below the cerebral cortex to broad areas of the central nervous system, such as the cerebral cortex, hippocampus, basal ganglia, cerebellum and spinal cord (Lungarella et al., 2003, Cox and Krichmar., 2009). The importance of the neuromodulatory system vastly outweighs the proportion of brain space it occupies (Cox and Krichmar., 2009), as they can they signal the occurrence of relevant stimuli or events (e.g. novel stimuli, painful stimuli, rewards) by modulating the neural activity and plasticity of a large number of neurons and synapses (Lungarella et al., 2003). Moreover, biological value systems act as a probabilistic reward system for reinforcing learning and behaviour.

The value system of a developmental/epigenetic robot is used as a source of intrinsic reinforcement for learning and motivation, signaling the occurrence of salient sensory inputs and qualitatively evaluating the effects of actions (Huang and Weng, 2002, Lungarella et al., 2003). Value systems not only introduce biases for learning, but also modulate it by qualitatively evaluating the consequences of particular action (Lungarella et al., 2003). While the value systems of a human adult is highly complex due to a lifetime of social and environmental influences, in experimental work current approaches to modeling value systems focus on particular aspects of a value system, such as novelty and curiosity, e.g. (Huang and Weng, 2002, Oudeyer et al., 2007). In this way an epigenetic robot is attracted to novel situations where learning progress can be maximised.

7. Conclusion

While the Chinese Room argues against Strong AI (i.e. computational intelligence will never have "understanding", "intentionality", or a "mind" in the *deep* sense of these words), it does not prove that building "intelligent" machines is impossible. On the contrary, the Chinese Room argument highlights one important aspect that is missing from computational models of the mind - the lack of value systems for qualitatively evaluating experience.

Despite symbol grounding being founded on the idea of ascribing "meaning" to a symbol system, little grounding related research addresses models of meaning *per se.* Most approaches to symbol grounding focus on meaning as a problem of reference. While reference is an important aspect of meaning, other aspects of meaning - most notably "value" - have been neglected. For example, Ziemke and Sharkey (2001) consider the "semiotic status" of artificial organisms, concluding that as artificial organisms "lack an intrinsic 'life task' this strongly questions the idea of 'first hand semantics' or 'content for the machine' in todays robotic systems". Thus, this paper argues that to build intelligent machines value systems need to be modelled - a critical aspect of epigenetic robotics.

References

Albus, J. and Barbera, A. (2005). RCS: A cognitive architecture for intelligent multi-agent systems. *Annual Reviews in Control*, 29(1):89–99.

Barsalou, L., Yeh, W., Luka, B., Olseth, K., Mix, K., and Wu, L. (1993). Concepts and meaning. In Beals, K., Cooke, G., Kathman, D., McCullough, K., Kita, S., and Testen, D., (Eds.), *Chicago Linguistics Society 29: Papers from the parasession on conceptual representations*, pages 23–61. University of Chicago: Chicago Linguistics Society.

Blank, D., Kumar, D., and Meeden, L. (2002). A developmental approach to intelligence. In Conlon, S. J., (Ed.), *Proceedings of the Thirteenth Annual Midwest Artificial Intelligence and Cognitive Science Society Conference.*

Cangelosi, A. (2006). The grounding and sharing of symbols. *Pragmatics and Cognition*, 14(2):275–285.

Chalmers, D. (1992). *The Symbolic and Connectionist Paradigms: Closing the Gap*, chapter Subsymbolic Computation and the Chinese Room. Lawrence Erlbaum.

Chella, A., Frixione, M., and Gaglio, S. (2003). Anchoring symbols to conceptual spaces: the case of dynamic scenarios. *Robotics and Autonomous Systems*, 43(2):175–188.

Christiansen, M. and Chater, N. (1993). Symbol grounding - the emperor's new theory of meaning? In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, pages 155–160.

Cisek, P. (1999). Beyond the computer metaphor: Behaviour as interaction. *Consciousness Studies*, 6(11-12):125–42.

Clark, H. and Schaefer, E. (1989). Contributing to discourse. *Cognitive Science*, 13:259–294.

Coradeschi, S. and Saffiotti, A. (2000). Anchoring symbols to sensor data: Preliminary report. In *AAAI/IAAI*, pages 129–135.

Coradeschi, S. and Saffiotti, A. (2003). An introduction to the anchoring problem. *Robotics and Autonomous Systems*, 43(2-3):85–96. Special issue on perceptual anchoring. Online at http://www.aass.oru.se/Agora/RAS02/.

Cox, B. and Krichmar., J. (2009). Neuromodulation as a robot controller: A brain inspired design strategy for controlling autonomous robots. *IEEE Robotics and Automation Magazine*, September.

Davidsson, P. (1993). Toward a general solution to the symbol grounding problem: Combining machine learning and computer vision. *AAAI Fall Symposium Series, Machine Learning in Computer Vision: What, Why and How?*, pages 157–161.

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought.* A Bradford Book, MIT, London.

Gärdenfors, P. (2003). *How Homo Became Sapiens: On the Evolution of Thinking.* Oxford University Press, London.

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

Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, (1):43–54.

Harnad, S. (1995). *The "artificial life" route to "artificial intelligence." Building Situated Embodied Agents*, chapter Grounding Symbolic Capacity in Robotic Capacity, pages 276–286. Lawrence Erlbaum, New Haven.

Harnad, S. (2003). The symbol grounding problem. In *Encyclopedia of Cognitive Science.* Nature Publishing Group/Macmillan, London.

Huang, X. and Weng, J. (2002). Novelty and reinforcement learning in the value system of developmental robots. In *Lund University Cognitive Studies*, pages 47–55.

Law, D. and Mikkaulainen, R. (1994). Grounding robotic control with genetic neural networks. *Tech. Rep., Univ. of Austin, Texas*, AI94223.

Lungarella, M., Metta, G., Pfeifer, R., and Sandini., G. (2003). Developmental robotics: a survey. *Connection Science*, 15(4):151–190.

MacLennan, B. (1993). Grounding analog computers. *Think*, 2(1):48–52.

Mayo, M. (2003). Symbol grounding and its implication for artificial intelligence. In *Twenty-Sixth Australian Computer Science Conference*, pages 55–60.

Nakisa, R. and Plunkett, K. (1998). Evolution of a rapidly learned representation for speech. *Language and Cognitive Processes*, 13(1):105–127.

Newell, A. and Simon, H. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the ACM*, 19, 3:113–126.

Oudeyer, P.-Y., Kaplan, F., and Hafner., V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286.

Peirce, C. S. (1931-1935). *Collected Papers of Charles Sanders Peirce*, volume 1-6. Harvard University Press, Cambridge, MA.

Pfeifer, R. and Verschure, P. (1995). *The Artificial Life Route to Artificial Intelligence*, chapter The Challenge of Autonomous Agents: Pitfalls and How to Avoid Them, pages 237–263. Lawrence Erlbaum, Hillsdale, New Jersey.

Prem, E. (1995). Dynamic symbol grounding, state construction and the problem of teleology. In Mira, J. and Sandoval, F., (Eds.), *From Natural to Artificial Neural Computation. International Workshop on Artificial Neural Networks. Proceedings*, pages 619–26. Springer-Verlag, Berlin, Germany.

Prince, C. (2001). Theory grounding in embodied artificially intelligent systems. In *Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, Lund, Sweden.

Putnam, H. (1973). Meaning and reference. *The Journal of Philosophy*, 70(19):699–711.

Riga, T., Cangelosi, A., and Greco, A. (2004). Symbol grounding transfer with hybrid self-organizing/supervised neural networks. In *IJCNN04 International Joint Conference on Neural Networks*, Budapest.

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

Sloman, A. and Croucher, M. (1981). Why robots will have emotions. In *In Proc 7th Int. Joint Conference on AI*, pages 197–202.

Steels, L. and Kaplan, F. (2002). *Linguistic evolution through language acquisition: formal and computational models*, chapter Bootstrapping Grounded Word Semantics, pages 53–74. Cambridge University Press.

Steels, L. and Vogt, P. (1997). Grounding adaptive language games in robotic agents. In Harvey, I. and Husbands, P., (Eds.), *Advances in Artificial Life. Proceedings of the Fourth European Conference on Artificial Life*, Cambridge, MA. MIT Press.

Swarup, S., Lakkaraju, K., Ray, S., and Gasser, L. (2006). Symbol grounding through cumulative learning. In et al., P. V., (Ed.), *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication*, pages 180–191. Springer.

Vogt, P. (2002). The physical symbol grounding problem. *Cognitive Systems Research Journal*, 3(3):429–457.

Williams, M.-A., McCarthy, J., Gärdenfors, P., Karol, A., and Stanton, C. (2005). A framework for evaluating groundedness of representations in systems: from brains in vats to mobile robots. In *IJCAI 2005*.

Ziemke, T. (1999). *Understanding Representation in the Cognitive Sciences*, chapter Rethinking Grounding, pages 177–190. Plenum Press, New York.

Ziemke, T. and Sharkey, N. E. (2001). A stroll through the worlds of robots and animals: Applying jakob von uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica*, 134(1-4):701–746.

Zlatev, J. (2001). A hierarchy of meaning systems based on value. In *LUND UNIVERSITY COGNITIVE SCIENCE*, page 2001.