Embodiment vs. Memetics: Does Language Need a Physical Plant?

Joanna J. Bryson (joanna@ai.mit.edu)
Computers and Cognition Group; Franklin W. Olin College of Engineering Needham MA 02492 USA

Abstract

Although the embodied approach to AI has lead to a large number of advances in the field, there has been no convincing demonstration of one of its earliest promises: that it would solve semantic grounding. In this article, I argue that although physical embodiment no doubt plays a major role in human intelligence, it may not be a necessary attribute for artificial agents capable of participating as equals in our linguistic culture.

Introduction

There is no doubt that embodiment is a key part of human and animal intelligence. Many of the behaviors attributed to intelligence are in fact a simple physical consequence of an animal's body (Raibert, 1986; Port and van Gelder, 1995). Taking a learning or planning perspective, the body can be considered as bias, constraint or a prior for both perception and action which facilitates an animal's search for appropriate behavior (Bryson, 2001).

This paper does not contest the importance of understanding embodiment to understanding human intelligence as a whole. This paper *does* contest one of the prominent claims of the embodied intelligence movement — that embodiment is the only means of 'grounding' semantics. One of the motivations for embodied AI has been the claim that a physical plant can solve the problem of semantics that has haunted artificial attempts to process and produce contentful natural language (NL). However, despite impressive advances in the state of artificial embodiment (e.g. Schaal, 1999; Kortenkamp et al., 1998), there have been no clear examples of artificial NL systems improved by embodiment.

I believe this is because embodiment is not a sufficient explanation of semantics, although we *have* seen some neat examples of the embodied acquisition of limited semantic systems (e.g Steels and Vogt, 1997; Steels and Kaplan, 1999; Billard and Dautenhahn, 2000). These systems show not only that semantics can be established between embodied agents, but also the relation between the developed lexicon and the agents' physical plants and perception. However, such examples give us little idea of how words like 'infinity', 'social' or 'represent' might be represented. Further, they do not show the *necessity* of physical embodiment for a human-like level of comprehension of natural language semantics. In contrast, it is possible that the semantic system underlying abstract words such as 'justice' may also be sufficient for terms originally referencing physical reality.

In this paper, I describe a disembodied model of the interaction between semantics and embodiment. I claim that, for humans, semantics is another form of automated perceptual learning, and is not *necessarily* related to our experience. We can acquire and transmit information without fully understanding$_{e/m}$ it[1]. By this I mean, we usefully transmit, and possibly even augment, information, while still failing to link the lexical entries we use to an underlying representation as rich or powerful as that which originally generated them. For example, if you were a physicist, I might say $E = MC^2$ to you and give you the idea for a massive energy source, while I myself never even read, let alone understood, the contents of Einstein's theory of special relativity. Under this theory, semantics in humans serves as one of many disjoint areas of expertise, which may or may not be linked to grounded$_{e/m}$, behavioral knowledge.

Some Definitions

Because we are in the process of trying to understand what 'semantics' and 'embodiment' mean, it follows that every paper will have slightly different connotations for these and related terms. This section specifies how I am using these terms in this paper. These are special usages and are flagged by the "Embodiment vs. Memetics" subscript. They are not currently the ordinary usages of the terms, nor do I mean to suggest they necessarily should be.

First, the basics:

- *semantics*$_{e/m}$: how a word is used.
- *plant*$_{e/m}$: any part of an agent that might directly impact or be impacted upon by the agent's environment.
- *expressed behavior*$_{e/m}$: behavior that impacts the environment, and is consequently externally observable.
- *grounded*$_{e/m}$: linked to, part of, or associated with a representation that determines an expressed behavior.

[1] Subscripts indicate specialized, paper-specific usages of terms, see definitions in Section

- *understand*$_{e/m}$: connect a semantic$_{e/m}$ term to a grounded$_{e/m}$ concept.

Embodiment$_{e/m}$ is having a plant. Notice that this means that software agents are embodied$_{e/m}$. I would argue that embodiment$_{e/m}$ is really a continuum: having more and richer interactions with a richer environment clearly increase potential for interesting grounding$_{e/m}$. Thus a VR agent with complex physics and actuators such as those created by Tu (1999) or Maes et al. (1994) might actually be more embodied than a mobile robot with only infrared sensors and a cylindric, limb-less plant$_{e/m}$.

Semantics without Reference

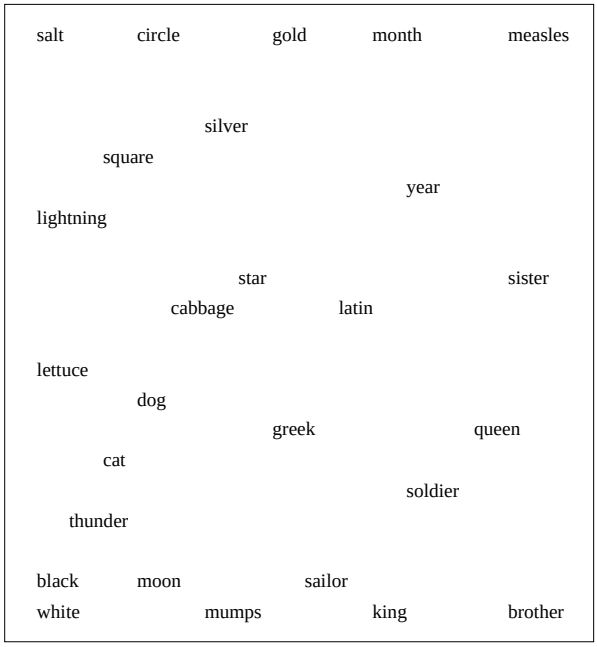

Figure 1: A two-dimensional projection of a semantic space, after Lowe (1997). The target words are taken from the experiments of Moss et al. (1995). Additional information on nearness is contained in the weights between locations in the 2-D space.

A basic premise of this paper is that human-like semantics$_{e/m}$ can be derived without any particular plant or embodiment$_{e/m}$. This isn't a hypothesis, it's a demonstrated fact. The demonstrations are in the computational psycholinguistics literature (McDonald and Lowe, 1998; Landauer and Dumais, 1997). This cognitive-science approach uses *semantic space* to define lexical semantics.

The underlying motivation for a semantic space model is the observation that *whatever* the reason that two words are similar in meaning, it shows up in the distributional profiles of the words — in how they are used, in the sorts of words they co-occur with. A semantic space associates each word with a vector of surrounding word co-occurrence counts so that distributionally similar words are associated with points nearby in the vector space. Distance or angular structure in the space represents semantic similarity highly effectively, as measured in a wide variety of experimental paradigms. In contrast to standard memory models, semantic space is also the only approach currently known that makes predictions about choice and reaction time behavior with real data derived purely from ambient language sources like newspaper text (see also Levy and Bullinaria, 2001; Finch, 1993).

The fact semantics can be acquired through purely statistical methods from everyday language should not be so surprising. We know, for example, that blind or paralyzed individuals typically learn to speak in a manner indistinguishable from able-bodied speakers, despite lacking many of the experiences a grounded semantics would seem to require (e.g. Landau and Gleitman, 1985). We do not assume that they do not, in fact, know what they are talking about when they use language, even language implying a physical ability that they lack (e.g. visual metaphors).

One interesting consequence of the semantic space model is that syntax is no longer a separate entity from semantics$_{e/m}$. Syntax is also a part of how language is used. Syntactic categories discriminate trivially in semantic space. *Starting* with syntax and trying to solve semantics later is barking up the wrong tree. This is analogous to the use of logic as a representation in AI. As MacDorman (1999) has argued, the use of overly-syntacticly based, semantically under-constrained representations is what has lead to the frame problem in AI (see also Harnad, 1993). In contrast, Gigerenzer and Todd (1999) demonstrate that humans and other animals use relatively constrained information in their reasoning.

I have been asked whether this sort of semantics$_{e/m}$ can lead not only to understanding, but to production. The short answer is "yes, of course". The long answer is that the question isn't even coherent within the framework described in the previous section. First of all, this semantics$_{e/m}$ does not necessarily lead to understanding$_{e/m}$, although it may do if at least some items of the lexicon are grounded$_{e/m}$. But secondly, I don't personally understand a concept of understanding$_{e/m}$ that does not directly impact expressed behavior$_{e/m}$. Statistical language models have already been used for generating fluent-sounding language (e.g. Oberlander and Brew, 2000), which is a large part of semantics$_{e/m}$ (see further Lapata et al., 2001). The real question is whether such a system can support semantic$_{e/m}$ word choice can be primed by the intentions of an artificial agent — a question I address somewhat below.

For the time being, I will only reverse the challenge. Consider the problem of speech recognition. The currently most successful methods in engineering and AI for recognizing and processing language make no use of referential information. Jelenik's famous observation "Anytime a linguist leaves the group the recognition rate

goes up" (Jurafsky and Martin, 2000, p. 191) leads to the fundamental problem of language modeling: since we 'know' that simple statistical time series models are incorrect models for language production and perception, how can we add more appropriate structure knowledge to our models and not do worse? Currently, adding any detailed referential knowledge or indeed to any other kind of grounding to speech recognition systems reliably worsens their performance. Until this problem is solved, what is the evidence that humans do use something as elaborate as embodied semantics?

The Semantic Species

Deacon (1997) proposes that the difference between humans and other animals is our ability to, having learned grounded$_{e/m}$ lexical concepts through experience, develop a web of relations between these concepts. The previous section shows us that in fact, humans could very well develop this web *independently* of the process of developing a grounded lexicon. I want to be clear here: in my model, humans still acquire associations between these two representations, just as in Deacon's model[2]. That's what I mean by understanding$_{e/m}$. The advantages of my model over Deacon's are: it better explains how abstract lexical terms are learned, it provides a common representation for all kinds of lexical semantics, and it allows for a rich model of insight and analogy, as new linkages can be formed between representations.

On the other hand, this model leads to a number of questions:

1. Precisely what does the association between semantics and grounding buy us?
2. To what extent does an artificial agent need grounding? Or in terms of the previous question, how *critical* is whatever grounding buys us?
3. If semantics is simple perceptual learning, why do only humans have such elaborate language and culture? This is actually a standard question about any language model.

These questions are addressed in the remainder of this section.

1. Representation without Intelligence

How much value is a system of knowledge a human might incorporate without understanding? On one hand, the answer is simple: how much use is a book? A book has no understanding of the information it transmits. Similarly, humans easily transmit information they don't understand, or don't understand fully. In fact, by mutation as well as recombination, humans may *generate* culture they don't fully understand. For example, there are several well-known anecdotes about scientists coming to a theory through misunderstanding a peer.

An interesting (or at least salacious) illustration of this point was the announcement by Juliana Hatfield during the publicity for her second solo alternative-rock album "Become What You Are" (Hatfield, 1993) that she was a virgin. This generated an enormous controversy in the music press, as people argued about the likely veracity of her statement. For evidence, they had (at minimum) two albums of music and lyrics by the artist concerned, the content of much of which related directly to the topic of relationships. Nevertheless, informed observers could not categorically determine whether she spoke the truth. This begs the question, how much embodied experience (that is, understanding$_{e/m}$) is necessary for the generation of salient culture[3]?

2. The Body Myth

Of course, there is an interesting relationship between semantic and grounded knowledge. We can, to some extent, deliberately guide our expressed behavior by rules we have learned using semantic terms. However, this process does not make us expert or graceful in a technique. Skilled behavior seems to derive from practice, or having pre-established skills that are readily applicable to a new ordering. Similarly, we can sometimes express verbally knowledge we have acquired non-verbally, by careful observation of our own behavior. However, this technique is again famously fragile, as we are often oblivious of essential steps in our processes, not to mention the difficulty of finding semantic terms for physical skills. Nevertheless, there can be no doubt that, despite their flaws, humans use these processes both to generate semantic content and cultural contributions, and to generate new expressed behavior.

However, this admission begs the question, does language *need* to be used this way? Hinton and Nowlan (1987) demonstrate in a model of the Baldwin Effect that evolutionary forces are unlikely to learn things that every individual agent reliably acquires on its own. Similarly, whatever we learn about language through grounding in our physical experience, could a society of agents without agents eventually acquire independently? Could we, in principle, build a 'Chinese Room' (Searle, 1980) that would participate in human conversation, and even generate new cultural content? This is an empirical question, but for the short term, my answer is "I don't see why not."

Of course, as I stated earlier, this would *not* be a complete model of human intelligence. Clearly humans have grounded aspects of their intelligence. Further, many of their expectations for social partners include physical plant$_{e/m}$ (Breazeal, 2000; French, 2000). However, we also willingly form at least casual relationships with people we have never communicated with in person (for example, chat room acquaintances), or who have different backgrounds or experiences (for example, people of different genders). We might extend this curtousy to ar-

[2]Though note that the grounded behavioral lexicon is almost certainly also modular, so there are probably more than two sets of representations becoming linked.

[3]A less salacious but perhaps deeper discussion of this question with respect to vision was produced by (Magee and Milligan, 1995).

tificial agents, particularly if they proved interesting conversational partners. More importantly, just as no model of human intelligence would be complete without embodiment, so also would it not be complete without including these disembodied linguistic capabilities.

3. Meme Machines and the Missing Link

Finally, if semantics is only a matter of perceiving statistical regularities, then why do humans appear to be the only species that rapidly accumulates culture? If social learning and cultural evolution are natural processes that happens relatively independently of slow, embodied learning, then why aren't they running in other species?

Recent work in primatology tells us three interesting things. First, we know that apes and even monkeys do have culture (de Waal and Johanowicz, 1993; Whiten et al., 1999). That is, behavior is reliably and consistently transmitted between individuals by non-genetic means. So we know that the question is not "why doesn't animal culture exist", but rather "why isn't it on the same scale as ours?"

Second, we know that primates have uniquely complicated social representations. For some time, this has been one of the basic hypotheses concerning why primates are so intelligent (Byrne and Whiten, 1988; Dunbar, 1995). But in particular, one of the critical aspects of Deacon's 'symbolic species' theory is that humans are particularly good at manipulating variable state — that we are somehow particularly good at having and redirecting deictic representations. On the other hand, I have already referenced arguments that over-generalized symbol manipulation is *not* a good basis for intelligence. Perhaps, the missing computational tool that primates have is the ability to represent relations *between* other agents. Harcourt (1992) presents evidence that all social species behave as if they keep record of relations between themselves and their group members (e.g. positive and negative interactions), but only primates behave as though they keep tabs on the relations between other agents. For example, apes will avoid fighting with close associates of dominant animals, and may try to befriend them (de Waal, 1996).

But if representing relations between conspecifics is the critical extra mental facility, and we share it with other primates, why don't other primates display explosive cultural growth? Perhaps there is another representation issue — this time the underlying representation which supports the disembodied communication of semantic content. If our memetic representation is a more fertile substrate for supporting unsupervised cultural evolution, then our culture would have a richer design space in which to evolve.

This leads to the third interesting discovery about primates: humans are the only species of primate capable of precise auditory replicative imitation (Fitch, 2000). My hypothesis is that the original basic unit of cultural transmission for humans was and is the auditory phrase. Auditory phrases are full of ordered information on a large number of axes: timing, duration, phonetics, and pitch.

There are a number of questions about this hypothesis, not least of which is whether other primates are capable of remembering precise timings for gestures: if not, they might have evolved a sign language as rich as our vocal one. However, if I am correct, and the trick is that the richness of the substrate representing the strictly semantic, ungrounded cultural transmission is the key, then we now have an explanation for why other primates don't share our level of culture. Birds have this same substrate (in fact, perhaps a richer one) but do not share the cognitive capacities of primates. The only other animals which might then hold a culture approximating our own are the cetaceans, the whales and dolphins. I will resist speculating about these animals.

	2^{nd}-ord. soc. rep.	no 2^{nd}-ord reps
vocal imit.	people	birds
no voc. imit.	other primates	most animals

Figure 2: Human-like cultural evolution might require both a rich memetic substrate such as provided by vocal imitation, and the capacity for second order social representations.

Contentions and Future Work

In this paper, I have described a rough model of human intelligence which includes both embodied, 'understood' knowledge and disembodied, memetic knowledge. While acknowledging that our culture would not be what it is today without embodied agents, I suggest that it now contains enough information that a disembodied agent might be able to gather sufficient semantic information directly from the culture to hold a conversation, or even to play a role in generating cultural content.

Of course, the model is far from complete — I have by no means specified all the interactions between these elements. But I propose that an interesting next phase of research is to build the 'Chinese Room' (in English) incorporating statistical natural language into an agent with basic provisions and motivations for turn taking, information seeking and knowledge sharing. We can then revisit the Turing test, and find out just how much more impressive such an agent is when its semantics are connected to an embodied, competent agent. Perhaps, with hard work and good engineering, we could connect this new English Room agent to a RoboCup player, and wind up with an artificial locker-room interview.

Acknowledgments

This paper wouldn't exist if not for many long discussions and several paragraphs from Will Lowe. Not that he necessarily agrees with me.

References

Billard, A. and Dautenhahn, K. (2000). Experiments in social robotics: grounding and use of communication in autonomous agents. *Adaptive Behavior*, 7(3/4).

Breazeal, C. (2000). *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. PhD thesis, MIT, Department of EECS.

Bryson, J. J. (2001). *Intelligence by Design: Principles of Modularity and Coordination for Engineering Complex Adaptive Agents*. PhD thesis, MIT, Department of EECS, Cambridge, MA.

Byrne, R. W. and Whiten, A., editors (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Oxford University Press.

de Waal, F. B. M. (1996). *Good Natured: The origins of right and wrong in humans and other animals*. Harvard University Press, Cambridge, MA.

de Waal, F. B. M. and Johanowicz, D. L. (1993). Modification of reconciliation behavior through social experience: An experiment with two macaque species. *Child Development*, 64:897–908.

Deacon, T. (1997). *The Symbolic Species: The coevolution of language and the human brain*. W. W. Norton & Company, New York.

Dunbar, R. I. M. (1995). Neocortex size and group size in primates: A test of the hypothesis. *Journal of Human Evolution*, 28:287–296.

Finch, S. (1993). *Finding Structure in Language*. PhD thesis, Centre for Cognitive Science, University of Edinburgh.

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, 4(7):258–267.

French, R. M. (2000). Peeking behind the screen: The unsuspected power of the standard turing test. *Journal of Experimental and Theoretical Artificial Intelligence*, 12:331–340.

Gigerenzer, G. and Todd, P. M., editors (1999). *Simple Heuristics that Make Us Smart*. Oxford University Press.

Harcourt, A. H. (1992). Coalitions and alliances: Are primates more complex than non-primates? In Harcourt, A. H. and de Waal, F. B. M., editors, *Coalitions and Alliances in Humans and Other Animals*, chapter 16, pages 445–472. Oxford.

Harnad, S. (1993). Problems, problems: The frame problem as a symptom of the symbol grounding problem. *Psycholoquy*, 4(34).

Hatfield, J. (1993). *Become What You Are*. Atlantic Records.

Hinton, G. E. and Nowlan, S. J. (1987). How learning can guide evolution. *Complex Systems*, 1:495–502.

Jurafsky, D. and Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, Englewood Cliffs, New Jersey.

Kortenkamp, D., Bonasso, R. P., and Murphy, R., editors (1998). *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*. MIT Press, Cambridge, MA.

Landau, B. and Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Harvard University Press, Cambridge, MA.

Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of induction and representation of knowledge. *Psychological Review*, 104:211–240.

Lapata, M., Keller, F., and Schulte im Walde, S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*, 30(4):419–435.

Levy, J. P. and Bullinaria, J. A. (2001). Learning lexical properties from word usage patterns. In French, R. M. and Sougné, J., editors, *Connectionist Models of Learning Development and Evolution: Proceedings of the* 6^{th} *Neural Computation and Psychology Works hop*. Springer.

Lowe, W. (1997). Semantic representation and priming in a self-organizing lexicon. In Bullinaria, J. A., Glasspool, D. W., and Houghton, G., editors, *Proceedings of the Fourth Neural Computation and Psychology Workshop: Connectionist Representations (NCPW4)*, pages 227–239, London. Springer-Verlag.

MacDorman, K. F. (1999). Grounding symbols through sensorimotor integration. *Journal of the Robotics Society of Japan*, 17(1).

Maes, P., Darrell, T., Blumberg, B., and Pentland, A. (1994). ALIVE: Artificial Life Interactive Video Environment. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, Seattle, Washington. AAAI Press.

Magee, B. and Milligan, M. (1995). *On Blindness : Letters Between Bryan Magee and Martin Milligan*. Oxford University Press.

McDonald, S. and Lowe, W. (1998). Modelling functional priming and the associative boost. In Gernsbacher, M. A. and Derry, S. D., editors, *Proceedings of the 20th Annual Meeting of the Cognitive Science Society*, pages 675–680, New Jersey. Lawrence Erlbaum Associates.

Moss, H. E., Ostrin, R. K., Tyler, L. K., and Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21:863–883.

Oberlander, J. and Brew, C. (2000). Stochastic text generatin. *Philosophical Transactions of the Royal Society of London, Series A*, 358.

Port, R. F. and van Gelder, T., editors (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press, Cambridge, MA.

Raibert, M. H. (1986). *Legged Robots That Balance*. mitpress, Cambridge, MA.

Schaal, S. (1999). Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, 3(6):233–242.

Searle, J. R. (1980). Minds, brains and programs. *Brain and Behavioral Sciences*, 3:417–424.

Steels, L. and Kaplan, F. (1999). Bootstrapping grounded word semantics. In Briscoe, T., editor, *Linguistic evolution through language acquisition: formal and computational models*. Cambridge University Press.

Steels, L. and Vogt, P. (1997). Grounding adaptive language games in robotic agents. In Husbands, C. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life (ECAL97)*, London. MIT Press.

Tu, X. (1999). *Artificial Animals for Computer Animation: Biomechanics, Locomotion, Perception and Behavior*. Springer.

Whiten, A., Goodall, J., McGew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E. G., Wrangham, R. W., and Boesch, C. (1999). Cultures in chimpanzees. *Nature*, 399:682–685.