Framing the debate

by

Glenberg, De Vega, Graesser

The purpose of this chapter is not to force a particular set of definitions or questions. Instead, we hope to lay out a set of definitions as a starting point. Other writers can either agree with these definitions or offer their own. Our goal is to be as precise as possible about what we mean when we use the following terms: symbol, embodiment, grounding. Following the proposed definitions are examples of the theories that are consistent with the definitions, and then a reiteration of the debate questions.

Symbols, grounding, and embodiment

C. S. Peirce's theory of signs is a good starting point both for how the theory helps and how it doesn't (see Clark, 1996 for a discussion of Peirce). According to Peirce, a sign is member of a triad consisting of an object, a sign, and an interpretant. An interpretant is in the mind of an observer; it is the idea of the object brought about by the sign. On this view, signs are a mechanism for grounding ideas, that is, the sign connects the idea to the object. Peirce proposed three types of signs. Icons connect the interpretant to the object by virtue of perceptual similarity. Photographs are prototypical icons. An index connects object and interpretant by virtue of a causal, spatial, or temporal relation. Thus, a weathervane is an index for the direction of the wind (a causal, spatial, and temporal connection), a pointing gesture is a spatial index for the pointed-to object, and the utterance of "I" is an index for the person speaking (a spatial and temporal connection). In contrast to signs that have a type of natural connection to their objects, symbols are signs only by virtue of a rule or convention. Thus, ordinary words are symbols. The word "chair" does not literally point to chairs or illustrate them in anyway. Furthermore, the conventional (or arbitrary) nature of a symbol is easily illustrated: The word that designates the object called a chair in English changes from language to language. Presumably, at least some icons and indices are less strongly tied to a language community (e.g., the meaning of a pointing gesture). Peirce also proposed that symbols (unlike icons and indices) are types. That is, a symbol by itself does not designate a particular thing, only types. Finally, Peirce noted that signs can be mixed: part symbol, part index, part icon.

For Peirce, signs (icons, indices, and symbols) are all mechanisms for grounding. That is, signs connect an idea (the interpretant) and the object. This notion is quite different from the use of the term symbol in much of cognitive science, and the use that is of concern here. In that use, symbols are taken to be a theoretical account of the interpretant, that is, a theoretical account of ideas. Nonetheless, cognitive scientists have adopted other aspects of the Peircean definition. Namely, symbols are types (they may be abstract), they may be arbitrarily related to objects, and they may be amodal, much like the word "chair" is the same symbol whether read or heard.

We offer one proposal for the definition of a symbol that has frequently been used in cognitive science. This definition has often been adopted by theoretical positions that emphasize the arbitrary relation between symbol and referent, as in the case of physical symbol systems (see Barsalou,1999; Harnad,1990; Newell,1980).

> A symbol is a theoretical element that is arbitrary, abstract, and amodal. Collections of symbols connected in the appropriate manner constitute ideas. Symbols and collections of symbols can be manipulated by using explicit rules to derive new ideas or to take action.

However, others in cognitive science have adopted a less constrained definition of symbol.. Researchers who develop symbolic models often speak of symbolic representations and procedures that have symbolic units directly linked to sensory transducers and motoric actuators (Just & Carpenter, 1992; Meyers, 2000; Kintsch, 1998; Kosslyn, 1994; Marr, 1982; Miller & Johnson-Laird, 1976; Norman & Rumelhart, 1975). Some symbols do have such links to modalities; others don't. Some symbols have merely conventional links to referents; others are more tightly grounded to perception and action. However, a symbolic position allows for the possibility of purely arbitrary links from symbolic code to sensori-motor referents and this opens the door wide open to an interesting debate with those who adopt a radical embodied position. In the interest of promoting clarity, therefore, it is necessary to qualify the term symbol with the type of symbol and its functional properties. There are amodal symbols, motoric symbols, perceptual symbols, indexical symbols, arbitrary symbols, and so on. Confusions are likely to develop without a more precise specification of the type of symbol and its properties

The term "representation" is also a construct that needs a precise definition, but there is the worry that proposing such a definition will open a Pandora's box of disagreements. Some view representations as a level of code that "stands for" the referent, whereas others consider representations has having constraints that are closely aligned with the psychological content or neurophysiological activities. Some researchers may take the position that it is best to not use the term "representation" in any explanatory sense because it creates confusion and may end up simply passing the buck. If a collection of symbols represents an idea, than what is the idea itself? We take it that one of the goals of theory is to explicate what an idea is, not to produce representations of ideas. Words and language seem to be perfectly good representations (signs for) ideas, so our question is how those words work. We believe that nothing is gained theoretically, or methodologically for that matter, by speaking of a representation without specifying more precisely what the nature of the representation is. In particular, for the purposes of this workshop, we need to know precisely the extent to which the representations are embodied and grounded in sensori-motor experience, as opposed to being abstract, amodal, and arbitrary..

With this initial definitions of a symbol, we can immediately start to frame questions for the debate. For example, can collections of amodal arbitrary symbols alone constitute ideas? Note that if the answer is "yes," then there is no reason to suspect that ideas are the sole province of biological systems. Why? Because abstract, amodal, and arbitrary symbols are the mechanism by which simple computers and abstract logical systems represent human ideas. If the answer is "no," then what else is needed to constitute ideas? We might need advanced computer architectures that capture the constraints of perception and action at a fine-grained level, as researchers have advocated and already developed in the areas of machine vision and robotics. Or we might need to consider the constraints of perception, action, and sensorimotor experience to be the fundamental guide in shaping our scientific understanding of language comprehension.

Very often the "what else" is framed in terms of *grounding*. A brief survey of sources that use the term "grounding" in their titles (Glenberg & Kaschak, 2002; Harnad, 1990; Pecher & Zwaan, 2005; Roy, 2005; Steels, 2003) does not reveal much in terms of definition. Here is what Roy (2005) says, and we will offer it as our definition of the process:

> The term grounding will be used to denote the processes by which an agent (human or machine) relates mental structures to external physical objects.[1]

Thus, we can ask a variety of questions about symbols. What sorts of things can be done with amodal arbitrary symbols that do not require grounding (e.g., syntactic transformations)? Can ideas composed of symbols be meaningful (be about something in particular, say, your favorite chair) without grounding? Do all symbols need to be grounded, or can some collections of symbols constitute meaningful ideas because they include some proportion of grounded symbols? What proportion is enough? Finally, when using rules to manipulate symbols, do the grounding processes necessarily play a role, or does grounding only matter for the initial induction of a symbol and the interpretation of the results of symbol manipulation?

Another answer to the "what else" question is embodiment. There have been several attempts to characterize the notion of embodiment, and a notable one is provided

[1] Note that this admirably clear definition seems to break down in various ways. For example, could a mentalistic term such as "idea" ever be grounded in the sense offered above. Probably a matter for the debate.

by Wilson (2002). Here we try for a different sort of explanation that might be useful for investigations primarily concerned with language and robotics. Before presenting that definition, however, a preamble is necessary. There are several ways of defining embodiment that do not seem to advance thinking. First, it is commonly accepted that knowledge about the world depends (at least in part) on perception and action systems that are parts of the body. Thus, claiming that a theoretical approach is embodied just because it proposes that knowledge derives from perception does not serve to discriminate between theories or provide an advance. Second, there has been a tremendous influence of neuroscience on theories of cognition, and it is commonly accepted that cognitive processes are reflected in neural processes. Thus, claiming that a theoretical approach is embodied because it invokes neural mechanisms does not provide an advance. Given this preamble, our definition of embodiment in regard to language is:

> Linguistic symbols are embodied to the extent that (a) the meaning of the symbol (the interpretant) to the agent depends on activity in systems also used for perception, action, and emotion, and (b) reasoning about meaning, including combinatorial processes of sentence understanding, requires use of those systems.

Two examples of symbolic theories

Landauer and Dumais's (1997) exposition of Latent Semantic Analysis (LSA) makes it clear that they consider it an amodal symbolic theory. The LSA theory begins by tracking the frequency of occurrence of about 60,000 words across some 30,000 texts, forming a matrix with words as rows, texts as columns, and frequencies as the cell entries. After some pre-processing, the matrix is submitted to a singular value decomposition to reduce the dimensionality. The result is a matrix in which the words are coded with values on about 300 dimensions (in contrast to the 30,000 texts). Landauer and Dumais state, "… we suppose that word meanings are represented as points (or vectors; later we use angles rather than vectors) in *k* dimensional space…" (page 215). Later in the paper they write, "Given the strong inductive possibilities inherent in the system of words itself, as the LSA results have shown, the vast majority of referential meaning may well be inferred from experience with words alone" (page 227). It should be noted that the sense of referential meaning they have in mind here is the generic referential meaning of a word, such the meaning of "bicycle" in different texts and contexts. They do not have in mind a particular indexical or deictic reference during the instantiation of a particular text or experience. Their central claim, which can be empirically tested, is that much of the generic referential meaning of a word W can be bootstrapped from a model that simply considers statistical properties of the family of other words that accompany word W, without having to consider the physical and social world that may accompany text. In this sense, the vast majority of meaning of a word W may well be inferred from experience with other symbols. The obvious implication of this claim is that experience with the world is not important for the "vast majority of referential meaning" of a word (generic meaning, not indexical or deictic meaning).

LSA vectors appear to be prototypical abstract, amodal, and arbitrary symbols, namely sets of numbers. But that level is at the mere *implementation* level, to use David Marr's terminology, and in this case the implementation level is statistical. The question arises as to what these vectors denote, if anything, at Marr's level of *computational* *theory* and the level of *representation and algorithm*. An answer to this question is not straightforward. Certainly it is the case that the sets of numbers are derived from perception (of the words). It is also the case that words do not appear together for arbitrary reasons; they appear together because the objects named appear together in the world. Nonetheless, all real perceptual information is stripped away before coding in LSA in that (a) words are descriptions not the objects themselves, and (b) it is co-occurrence frequencies of words that matter. That is, frequency is the currency, not say, redness, or loudness, or spatial extent.

Landauer and Dumais (1997) do recognize the need for some sort of symbol grounding, “But still, to be more than an abstract system like mathematics words must touch reality at least occasionally” (page 227). To ground a word such as “rabbit” they suggest, “judiciously add[ing] numerous pictures of scenes with and without rabbits to the context columns in the encyclopedia corpus matrix, and fill[ing] in a handful of appropriate cells in the *rabbit* and *hare* word rows.” This would seem to potentially ground (potentially, because there are enormous practical problems with the suggestion of adding pictures) the LSA symbols in visual information. We don’t wish to pre-judge whether this would actually lead to grounding; that should be a matter of debate. We do wish to note, however, that the suggestion for adding pictures falls far short of creating an embodied theory as defined above. First, there is no need for activity in anything like a neural system because the meaning depends on frequencies of co-occurrences, not anything intrinsic to, say, visual properties or auditory properties or action properties. Second, reasoning in LSA is based solely on the frequencies (or the corresponding dimensional values), not directly on anything to do with perceptual or action systems.

For example, judging that a rabbit and a hare are similar does not require accessing visual information, instead, the judgment is based on a mathematical comparison of the similarity of the vector representations. The values of these vector representations are loosely constrained by the perceptual world, so there may be an indirect link with perception and action.

A second example of a symbolic theory (or implementation, see below) is provided by Rogers, Lambon Ralph, Garrard, Bozeat, McClelland, Hodges, and Patterson (2004). This example is interesting because whether one concludes that the theory is embodied or symbolic depends on whether one analyzes the verbal statement of the theory or the implementation as a connectionist network.[2] They state on page 206:

> …we suggest that the representations and processes underlying semantic memory are best understood within a theory in which semantic knowledge emerges from the interactive activation of modality-specific perceptual representations of objects and statements

---

[2] Is a connectionist theory an appropriate example of a symbolic theory? First, connectionist models are usually considered –by connectionists and symbolists alike- as something different from the symbolic approach. Thus, in many PDP models there are not individual symbols for particular concepts. Second, any attempt to implement a theory –included an embodied theory- forces one to fit the constraints of current computer and programming technologies: for instance, even connectionist models are a simulation of parallel processing in a serial computer. Third, computations in the brain could be ultimately described as patterns of activation in neural networks, such as in Pulvermuller's neo-Hebbian approach, which is not far away from connectionist intuitions.

Nonetheless, there may be several reasons for treating some types of connectionists accounts as symbolic. First, as in Rogers et al., some encodings may be localist in that one node corresponds to one idea, which is very much in conformation to our definition of a symbol. Second, as illustrated shortly, in some connectionist accounts, the manipulations are based only on frequency of occurrence or co-occurrence irrespective of the putative meaning or perceptual system.

> about objects. In contrast to some contemporary approaches, we argue that semantic representations do not need to extract, sort, and retrieve attributes, facts, or propositions about objects to fulfill this role; they need only to allow such information to be produced as overt responses in particular task contexts…The content of semantic memory is represented in the same regions of cortex that directly encode modality-specific regularities in the environment during perception and action…These acquired [semantic] representations do not code explicit semantic content, but they are structured in ways that facilitate the system's ability to generate appropriate responses when given perceptual inputs.

Later in the paper they write, "Semantic knowledge may be construed as a process that mediates the interactions among content-bearing perceptual representations, rather than as a repository of propositional facts about objects." One could hardly ask for a better match with our definition of embodiment.

The commitment to ideas of embodiment disappears, however, in the implementation of the model. There are three subnetworks consisting of localist encoding of (a) 64 visual features, (b) 152 verbal features (object names, "perceptual" features such color names, functional features such as "can fly', etc.), and (c) a 64 unit distributed, hidden-node semantic network . The model is trained so that the input of visual features, combined with the information in the semantic units, produces activation of appropriate verbal features, and input of verbal features, combined with the information in the semantic units, produces appropriate activation of visual features.

Note, however, that the "visual" features are nothing of the sort. Each feature is an on/off activation of a node. The fact that one node may be designated as the color red, and another node designated as "leg" and a third designated as "big" is irrelevant to the processing of the model. The only thing that counts is frequency, or better yet, co-occurrence frequency (coherent covariation to use Rogers et al's terminology). In fact, if there was another subunit coding auditory features, exactly the same processing rules would apply. The fact that one node would designate "loud" and another "high pitched" and another the phoneme /k/ is irrelevant. The activations on those nodes would be treated identically, and identical to the activation on the "visual" nodes. The point is that in the Rogers et al. implementation, there is nothing visual or verbal about the subnetworks; the only thing that matters is co-occurrence. Similarly, in terms of reasoning about meaning, the only information that matters is co-occurrence frequency; it is irrelevant that one node is labeled "leg" and another "loud." This conclusion is stated clearly in the Rogers and McClelland's (2004) exposition of the theory, "What is important to the model's behavior is the propensity for various sets of properties to covary across different items and contexts, regardless of whether the properties are conveyed through spoken statements or through other perceptual information. *It is not the identity of the properties themselves, but their patterns of covariation that is essential to the model's behavior*" (page 117, emphasis added).

It could be argued, however, that the model by Rogers and colleagues is in fact embodied at the level of combining the primitive units in the execution of action, but not at the level of identification and activation of the primitive units themselves. All models need to precisely declare what level of behavior or cognition is being explained versus what levels are assumed as primitive units. It is extremely important, as we debate the issues, that we specify precisely what phenomena we are attempting to explain as a direct object of inquiry, versus what phenomena and components are merely offered as a simplification in our working assumptions.

Two examples of embodied theories

The first example of an embodied theory is Barsalou's (1999) Perceptual Symbol System. We won't describe the model in detail. Instead we note that Barsalou describes a perceptual symbol for an object as based on the neural activity in multiple perceptual systems when the object was initially perceived. Furthermore, manipulation of knowledge requires a simulation using those same perceptual systems. Perhaps because this theory is not implemented, it can remain true to its embodied roots.

The second example of an embodied system is Roy's (2005) robot, Ripley, "a robotic manipulator that is able to translate spoken commands such as 'hand me the blue one on your right' into situated action." According to Roy (2005),

> Ripley's representations and algorithms led to an approach that grounds the meaning of verbs, adjectives, and nouns referring to physical referents using a unified representational framework…Verbs are grounded in sensory-motor control programs similar to x-schemas. Adjectives describing object properties are grounded in sensory expectations relative to specific actions…For example, the meaning of 'red' is not simply a color category, but rather a color category linked to the motor program for directing active gaze towards an object. 'Heavy' is grounded in haptic expectations associated with lifting

> actions. In this way, all perceptual properties are related to appropriate actions. Locations are encoded in terms of body-relative coordinates. Objects are represented as bundles of properties tied to a particular location along with encodings of motor affordances for affecting the future location of the bundle."

Does the implemented Ripley really encode the environment using its perception and effector systems, or does the commitment to embodiment break down as in Rogers et al (2004)? Fortunately, we will be able to pin Deb down at the workshop!

Given these proposed definitions and examples, many of the questions we formulated for the debate are still operative. Thus, in terms of the need for symbol grounding, is it the case that:

1. Searle's Chinese Room Argument [F3] is wrong: symbol systems need not be grounded to represent or produce meaning.
2. Language needs to be grounded, but the mechanism of grounding plays little role in most language processing.
3. Searle's argument is no longer relevant to contemporary symbolic models that are complex, multilayered at coarse and fine grain levels, and grounded in perception and action. Searle's argument is either incorrect or indeterminate.
4. Grounding language in embodied representations is important for understanding some aspects of language (e.g., about concrete situations), but not others (e.g., descriptions of some very abstract situations).

5. All levels of language understanding are grounded in action and perception.

In regard to the interpretation of neuroscientific data, is it the case that:

1. Word meaning activates brain regions that partially overlap those responsible of perception and action. Therefore, the embodiment of meaning has been empirically demonstrated.

2. Neurological data do not solve the question because, even if the processing of word meaning overlaps sensory-motor areas in the brain, this fact does not preclude that the processes themselves are symbolic.

3. In addition to the above claim, the classical perisylvian areas of language in the left brain hemisphere perform computations that are more symbolic-like (morpho-syntactical) than embodied.

In regard to whether computers can model comprehension that is embodied and symbolically grounded, which of the following can be most persuasively defended?

1. Because computers are not biological systems, they cannot be embodied or even simulate embodied cognition. Nonetheless, computers might be able to model embodied cognition in the same sense that a computer can model a thunderstorm by solving complex equations.

2. A computer program can simulate embodied cognition in humans, but cannot be literally embodied in the same way that human meaning is embodied.

3. A computer program can be embodied, but the nature of the embodiment is different from humans because the world experienced by the computer is different.

4. A computer/robot can be fully embodied in the same way that meaning is embodied in humans.

When considering Ripley in the context of our definition of embodiment, it would seem that the answer must be 3 or 4. However, this may indeed be challenged by some researchers who question the adequacy of Ripley in truly capturing the computational theory, representations, algorithms, and neuroscience implementation of an embodied theory.

Of course, all of this is just wasted words unless we figure out how to decide these issues empirically. Thus, we encourage you all to address in your chapters how these questions can be resolved, and to include your own data that contribute to such a resolution. In addition to the debate questions, we encourage you to tackle the following:

1. Describe an ideal experiment that would convince a symbolist that embodied representations are routinely activated to understand language.

2. Describe an ideal experiment that would convince an embody theorist that symbolic representations are necessary to understand at lest some kinds of linguistic expressions.

References

Barsalou, L. (1999). Perceptual symbols systems. *Behavioral and Brain Sciences*, 22, 577-609.

Clark, H. H. (1996). Using Language. New York: Cambridge University Press.

Newell, A. (1980). Physical symbol Systems. *Cogntive Science*, 4, 135-183.

Glenberg A.M., Kaschak M.P. (2002). Grounding language in action. Psychonomic Bulletin & Review, 9, 558-565

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, pp 335-346.

Just M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211- 240.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive control processes and human multiple-task performance: Part 1. Basic Mechanisms. *Psychological Review, 104,* 3-65.

Miller, G.A., & Johnson-Laird, P.N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Normal, D.A., & Rumelhart, D.E. (1975). Explorations in cognition. San Francisco: Freeman.

Pecher, D., & Zwaan, R. (2005). *Grounding Language*. New York: Cambridge University Press.

Rogers, T.T., Lambon Ralph, M. A., & Garrod, P. Bozeat, S, McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, 111, 205-235.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic Cognition*. cambridge, MA: The MIT Press:

Roy, D. (2005). Grounding words in perception and action: computational insights *Trends in Cognitive Sciences*, 9, 389-396.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cogntive Sciences*, 7, 308-312.

Wilson, M. (2002). Six views of embodied cognition. Psychonomic Bulletin & Review, 9, 625-636.