Journal of Memory and Language **43,** 379–401 (2000)
doi:10.1006/jmla.2000.2714, available online at http://www.idealibrary.com on IDEAL®

Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning

Arthur M. Glenberg and David A. Robertson

University of Wisconsin—Madison

Latent Semantic Analysis (Landauer & Dumais, 1997) and Hyperspace Analogue to Language (Burgess & Lund, 1997) model meaning as the relations among abstract symbols that are arbitrarily related to what they signify. These symbols are ungrounded in that they are not tied to perceptual experience or action. Because the symbols are ungrounded, they cannot, in principle, capture the meaning of novel situations. In contrast, participants in three experiments found it trivially easy to discriminate between descriptions of sensible novel situations (e.g., using a newspaper to protect one's face from the wind) and nonsense novel situations (e.g., using a matchbook to protect one's face from the wind). These results support the Indexical Hypothesis that the meaning of a sentence is constructed by (a) indexing words and phrases to real objects or perceptual, analog symbols; (b) deriving affordances from the objects and symbols; and (c) meshing the affordances under the guidance of syntax. © 2000 Academic Press

Key Words: meaning; language; embodiment; computational models; Latent Semantic Analysis; Hyperspace Analogue to Language.

Meaning is the most important problem in cognitive psychology. Meaning controls memory and perception. Meaning is the goal of communication. Meaning underlies social activities and culture: To a great degree, what distinguishes human cultures are the meanings they give to natural phenomena, artifacts, and human relations. Yet, rather than being a hotbed of theoretical and empirical investigation, meaning in cognitive psychology has been coopted by a particular approach: Meaning arises from the syntactic combination of abstract, amodal symbols that are arbitrarily related to what they signify.

A new form of the abstract symbol approach to meaning affords the opportunity to examine its adequacy as a psychological theory of meaning. This new form is represented by two theories of linguistic meaning (that is, the meaning of words, sentences, and discourses), both of which take advantage of the mathematics of high-dimensional spaces. The Hyperspace Analogue to Language (HAL; Burgess & Lund, 1997) posits that the meaning of a word is its vector representation in a space based on 140,000 word–word co-occurrences. Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) posits that the meaning of a word is its vector representation in a space with approximately 300 dimensions derived from a space with many more dimensions. The vector elements found in both theories are just the sort of abstract features that are prototypical in the cognitive psychology of meaning. Our goals are to investigate the adequacy of these high-dimensional theories both analytically and empirically and to contrast these theories of meaning

This work was supported in part by a grant to Arthur Glenberg from the University of Wisconsin—Madison Graduate School Research Committee, Project 990288, and a National Institute of Mental Health (T32-MH18931) predoctoral traineeship to David Robertson. This work benefited substantially from contributions of the students in the Honors Seminar in Cognitive Psychology (1997/1998) who helped with the design of the experiments, construction of the stimuli, and data collection. Those students were Christopher Amadon, Brianna Benjamin, Jennifer Dolland, Jeanette Hegyi, Katherine Kortenkamp, Erik Kraft, Nathan Pruitt, Dana Scherr, Sara Steinberg, and Brad Thiel. We thank Curt Burgess and two anonymous reviewers for comments on the manuscript.

Address correspondence and reprint requests to Arthur M. Glenberg, Department of Psychology, 1202 West Johnson Street, Madison, WI 53706. E-mail: glenberg@facstaff.wisc.edu.

0749-596X/00 $35.00

with a theory that does not make use of abstract, arbitrary symbols.

The structure of this article is as follows. First, we describe in more detail the HAL and LSA models and the associated claims that they are adequate theories of meaning. Second, we introduce the symbol grounding problem (Harnad, 1990) and review why it is a touchstone for determining the adequacy of any theory of meaning. Third, we discuss how the high-dimensional theories attempt to solve the symbol grounding problem. Fourth, we sketch an alternative account of meaning and an alternative solution to the symbol grounding problem based on an embodied theory of cognition (Glenberg, 1997). Fifth, we present three experiments designed to test various solutions to the symbol grounding problem. The experiments reveal serious shortcomings in the high-dimensional theories, but the results of the experiments are generally in accord with predictions from the embodied theory.

The HAL theory has been described in several publications; here we use Burgess and Lund (1997) as our main reference. In the HAL theory, word meaning is derived from a dimensional analysis of words in context. A large corpus (e.g., 300 million words) is analyzed using a moving window of 10 words and a matrix of some 70,000 rows and columns. Each row and column is labeled with a particular word so that the cells of the matrix can be used to record the co-occurrence of pairs of words. As a word is encountered, co-occurrence values are added to the matrix to represent the closeness of the encountered word to other words in the 10-word window. Words that are adjacent are given a co-occurrence value of 10; those separated by one other word are given a value of 9; and so on. At the end of the process, rows give the total co-occurrence values for words which precede the row label, whereas columns give the co-occurrence values for words following the column label. Then, for a given word, the 70,000 element row vector and the 70,000 element column vector are conjoined to produce a 140,000 element vector: HAL's proposed meaning of the word. An important aspect of the co-occurrence data extracted by HAL is the similarities of contexts in which words occur. For example, STREET and ROAD are coded as similar by HAL not because they appear frequently in the same sentences, but rather because the 10-word windows for street are similar to the 10-word windows for road.[1]

In what way can this vector be a representation of meaning? As Burgess and Lund (1997) note, the vector representations correlate with human performance in tasks thought to tap meaning. For example, Lund, Burgess, and Atchley (1995) demonstrated that similarity among vectors correlated with degree of priming in a lexical decision task. Burgess and Lund (1997) used the vectors to simulate categorization. That is, they identified words in several categories (animal types, body parts, cities, and geographical locations) and submitted the corresponding vectors to a multidimensional scaling (MDS) procedure. The scaling procedure roughly grouped the vectors into their categories, although there were some notable misgroupings. For example, the vectors for "finger" and "leg" were closer in the MDS space to "cat" and "mouse" than to "hand" or "foot."

Landauer and Dumais (1997) offer LSA as a theory of acquisition, induction, and representation of knowledge. In this theory, words are also represented as vectors derived from co-occurrence in text, but there are a number of differences between LSA and HAL. In deriving the LSA vectors, one first selects a "semantic space"; that is, a set of contexts. Landauer and Dumais (1997) describe one space as consisting of the first 2000 characters in each of 30,473 articles in the electronic version of Grolier's Academic American Encyclopedia. Each of the 30,473 articles is assigned a column in the matrix, and each of the 60,000 some words is assigned a row. The entries in the matrix are the number of times in which a word occurs in a context (article). These entries are logarithmi-

[1] As a shorthand description, we refer to HAL and LSA matrices and vectors as coding co-occurrence. Nonetheless, local word-to-word co-occurrence may be of less importance than global co-occurrence, which is "the weighted collection of local co-occurrences or the context history of a word" (quoted from Curt Burgess's review of the manuscript upon which is article is based).

cally transformed and subjected to a singular value decomposition (SVD). SVD is similar in effect to factor analysis or the computation of eigenvalues. The result of the SVD analysis is the extraction of about 300–400 important dimensions and each word's values on the dimensions. Thus, each word is eventually represented as a vector of about 300–400 numbers.

Landauer and Dumais make a number of claims for LSA vectors. As examples, they note that a radical interpretation of their results (which does not seem to be disavowed by the authors), takes LSA "as a possible theory about all human knowledge acquisition, as a homologue of an important underlying mechanism of human cognition in general" (p. 212). More specifically, "we suppose that word meanings are represented as points . . . in k dimensional space . . ." (p. 215) and that an appropriate set of dimensions may be considered "a unified representation of knowledge" (p. 217). Although Landauer and Dumais discuss the symbol grounding problem (reviewed later in this article), they also suggest that "the LSA results have shown [that] the vast majority of referential meaning may well be inferred from experience with words alone" (p. 227); that is, from noting which words occur in which contexts.

Landauer and Dumais also apply LSA to sentence and discourse understanding. A sentence is represented as the average of the vectors of the words it contains, and the coherence between sentences is predicted by the cosine of the angle (in multidimensional space) between the vectors corresponding to successive sentences. They claim that LSA averaged vectors capture "the central meaning" of passages (p. 231).

Like Burgess and Lund, Landauer and Dumais offer a variety of demonstrations in support of these claims. First, LSA can retrieve documents that are meaningfully related to queries that do not contain the same words as the documents (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). Second, LSA can mimic performance of nonnative English speakers who take the Test of English as a Foreign Language. That is, LSA vectors can pick out synonyms about as effectively as the nonnative English speakers (Landauer & Dumais, 1994). Third, LSA's rate of growth of knowledge (per paragraph of text) is comparable to some measures of knowledge growth in children (Landauer & Dumais, 1997). Fourth, LSA average sentence vectors can predict coherence judgments (Landauer & Dumais, 1997). Fifth, LSA vectors can be used to score essays (Landauer, Laham, Rehder, & Schreiner, 1998), although the procedure appears to depend on first having humans score a subset of the essays.

HAL and LSA have achieved an impressive number of demonstrations of their predictive (or correlational) validity. But, can they be theories of human meaning? It is not controversial that a word (i.e., a sequence of letters or sounds) is nothing more than an abstract symbol arbitrarily related to its referent (what the word refers to), and thus each word is in need of some sort of definition or grounding. Both HAL and LSA appear to be proposing that the meaning of an abstract symbol (a word) can arise from the conjunction of relations to other undefined abstract symbols; that is, that meaning arises from the conjunction of relations implicit in the HAL and LSA matrices. Will this work? Searle's (1980) Chinese Room Argument is meant to demonstrate that abstract, arbitrary symbols, such as words, need to be grounded in something other than relations to more abstract arbitrary symbols if any of those symbols are to be meaningful.

Consider a thought experiment (adapted from Harnad, 1990, and related to the Chinese Room Argument) that suggests that something critical is missing from HAL and LSA. Imagine that you just landed at an airport in a foreign country and that you do not speak the local language. As you disembark, you notice a sign printed in the foreign language (whose words are arbitrary abstract symbols to you). Your only resource is a dictionary printed in that language; that is, the dictionary consists of other arbitrary abstract symbols. You use the dictionary to look up the first word in the sign, but you don't know the meaning of any of the words in the definition. So, you look up the first word in the definition, but you don't know the meaning of the words in that definition, and so on. Obviously, no matter how many words you look up, that is, no matter how many structural relations you determine among the arbitrary abstract symbols, you will never figure out the meaning of any of the words. This is the symbol grounding problem (Harnad, 1990): To know the meaning of an abstract symbol such as an LSA vector or an English word, the symbol has to be grounded in something other than more abstract symbols.

Consider three solutions to the symbol grounding problem. The first is to depend on perception. On this solution, part of language acquisition requires instruction in mapping of cognitive abstract arbitrary symbols to specific referents. Whereas that is easy to say, there have been no demonstrations of how this might be done (Barsalou, 1999). The problem is daunting: How does stimulation in all of its sensory richness and situational particularities become stripped down to an arbitrary, amodal, abstract symbol? There are several logical problems, also. Most abstract symbol systems are Aristotelian: a symbol is either assigned to a perception or not (in contrast to fuzzy systems, Oden, 1984, 1987). However, to a great degree the natural world is not Aristotelian but continuous and overlapping. The symbol grounding problem is just as daunting in going the other direction, from thought to the world. We can imagine a symbol manipulation system deriving relations among symbols through logical or computation processes; that is, like LSA determines similarity among vectors by computing the cosine between the vectors. But how does the system know what the symbols are about; that is, what it is thinking about? Again, the solution appears to be to map the cognitive symbols onto their real-world referents. Putnam (as discussed in Lakoff, 1987) has demonstrated the impossibility of this solution. In brief, Putnam demonstrates how any system of relations among abstract symbols can be mapped onto a great variety of different real-world referents having the same relations (just as a set of mathematical equations can be mapped onto a great variety of real-world constructs). Thus, there is no way for the system to know with certainty what it is thinking about.

A second solution to the symbol grounding problem is offered by both Landauer and Dumais (1997) and Burgess and Lund (1997). Landauer and Dumais summarize the symbol grounding problem by noting, “But still, to be more than an abstract system like mathematics words must touch reality at least occasionally” (p. 227). Their proposed solution is to encode, along with the word stream, the streams from other sensory modalities. “Because, purely at the word–word level, rabbit has been indirectly preestablished to be something like dog, animal, object, furry, cute, fast, ears, etc., it is much less mysterious that a few contiguous pairings of the word with scenes including the thing itself can teach the proper correspondences. Indeed, if one judiciously added numerous pictures of scenes with and without rabbits to the context columns in the encyclopedia corpus matrix, and filled in a handful of appropriate cells in the rabbit and hare word rows, LSA could easily learn that the words rabbit and hare go with pictures containing rabbits and not to ones without, and so forth” (p. 227). Burgess and Lund (1997) offer a similar solution, “We do think a HAL-like model that was sensitive to the same co-occurrences in the natural environment as a human-language learner (not just the language stream) would be able to capitalize on this additional information and construct more meaningful representations” (p. 29).

There are several reasons to question this solution. First, the solution is not implemented in either LSA or HAL. Second, the solution appears to presuppose that the symbol grounding problem has been solved. For example, to implement the procedure Landauer and Dumais suggest, the LSA program would need to know which pictures (or real-life scenes) actually contained rabbits and which did not in order to add the pictures as separate columns to the matrix. Furthermore, the symbol grounding problem needs to be solved for the program to know which rows in the matrix (words) should be incremented for each picture. Third, there are empirical grounds for questioning this solution. Suppose that a child’s sole exposure to language was through watching television. The audio channel contains information similar to the word stream that LSA is exposed to. The video channel implements a second sensory modality, as suggested by the proposed solution to the symbol grounding problem. However, both Pinker (1994) and Ervin-Tripp (1973) discuss data indicating that children cannot learn a language solely from watching TV. Finally, as we demonstrate below, even if all of these problems could be solved, the sort of information encoded in HAL and LSA vectors appears to be, in principle, insufficient for ordinary language understanding.

The third solution to the symbol grounding problem, but one that is incompatible with both LSA and HAL, is to drop the assumption that meaning is based on abstract symbols arbitrarily related to their referents. This type of solution is being used by an increasing number of researchers (e.g., Barsalou, 1999; Glenberg, 1997; Lakoff, 1987; MacWhinney, 1998; Newton, 1996) investigating meaning from the perspective of embodied cognition. Here we sketch one such solution based on Glenberg (1997; Glenberg & Robertson, 1999) that provides an alternative to LSA and HAL. This alternative is contrasted with LSA and HAL in the following experiments.

Glenberg (1997) proposed that cognition evolved to coordinate effective action; that is, action that enhances survival and reproductive success given the constraints of a particular type of body. The structure of the body is enormously important in choosing effective action. For example, when faced with a dangerous situation, effective actions for a mole, a bird, and a human are quite different. Given the importance of action for survival, and given that meaning of a situation is a cognitive construal, Glenberg (1997) suggested that the meaning of a particular situation for a particular animal is the meshed (i.e., coordinated) set of actions available to that animal in that situation.

This set of actions depends on several components. First, the set of actions (and thus meaning) depends on the affordances (Gibson, 1979) of the situation.[2] Affordances are based on the relation between objects and bodily abilities. For example, a chair affords sitting to beings with humanlike bodies, but it does not afford sitting for elephants. A chair also affords protection against snarling dogs for an adult capable of lifting the chair into a defensive position, but not for a small child. Second, the set of actions depends on the individual's learning history, including personal experiences of actions and learned cultural norms for acting. Thus, a chair on display in a museum affords sitting, but that action is blocked by cultural norms. Third, the set of actions depends on the individual's goals for action. A chair can be used to support the body when resting is the goal, and it can be used to raise the body when changing a light bulb is the goal.

These determinants of action are meshed to form a coordinated set of actions. Meshing affordances, experiences, and goals requires that the various types of actions be integrated in a manner that respects intrinsic constraints on bodily activity that arise from biology and physics. That is, in a real body not all actions can be combined. For example, a real human body cannot simultaneously sit and jump, although it can sit and eat or sit and swing its legs. The various components of meaning (affordances, experiences, and goals) can be meshed because they are all realized in the domain of bodily activity rather than in abstract, amodal, arbitrary representations. When affordances, experiences, and goals are successfully meshed, they form a coherent, doable, and envisionable set of actions: the individual's meaningful construal of the situation.

Glenberg and Robertson (1999) developed the Indexical Hypothesis to relate the general theory of embodied cognition to language comprehension. According to this hypothesis, understanding a sentence such as "Jareb stood on the chair to change the light bulb" requires three processes. The first is to index phrases to actual objects or analogical perceptual symbols (Barsalou, 1999) representing the objects. Thus, the noun phrase "the chair" may be taken to refer to an actual chair in the perceiver's environment or indexed to a prototypical representation of a chair that retains perceptual information; that is, a perceptual symbol. Barsalou (1999), Barsalou,

[2] Unlike Gibson, however, we do not claim that affordances are necessarily directly perceived. At the very least, learning can lead to the detection of new affordances, as discussed later.

Solomon and Wu (1999), and Horton (1998) present evidence demonstrating the reality of perceptual symbols. The second step is to use the indexed object or perceptual symbol to derive affordances. The third step is to mesh the affordances guided by the syntax of the sentence (Kaschak & Glenberg, 2000). Thus, one meshes the affordances of the chair and of Jareb so that Jareb is on the chair rather than, say, under the chair. A sentence is meaningful to a particular reader to the extent that the reader can mesh the objects and activities as directed by the sentence. If for a particular reader "Jareb" was a pet fish or a baby, the sentence would not make sense because chairs do not afford standing-on for fish or babies.

As another example, consider the difference between sentence (1a) and (1b).

(1a) After wading barefoot in the lake, Erik used his shirt to dry his feet.
(1b) After wading barefoot in the lake, Erik used his glasses to dry his feet.

Sentence (1a) makes sense, but sentence (1b) does not. Why not? Sentence (1b) is grammatical, propositions can be formed and embedded, and as we demonstrate below, the associative relation between "glasses" and "dry" is just as strong (or weak) as that between "shirt" and "dry." The reason sentence (1b) does not make sense is because the affordances of glasses do not mesh with the actions required to dry one's feet. On the other hand, it is trivial to make sentence (1b) sensible by simply changing the affordances of the referent of "glasses." So, if we learned that Erik was clowning around and wearing a large pair of glasses carved out of a sponge, we can envision the mesh of glasses and drying. The point is that understanding the sentence requires knowing the affordances of the referents.

The indexical hypothesis and high-dimensional theories of meaning (LSA and HAL) differ in several ways. First, the cognitive elements of the indexical hypothesis (e.g., affordances and perceptual symbols) are not arbitrarily related to what they represent. Affordances are directly related to the interaction of bodily capabilities and the situation. Thus, part of the symbol-grounding problem is eliminated. One need not search for the mapping between abstract arbitrary symbols and objects in the world. Instead, the cognitive elements can be matched to perceptual experience. Second, understanding relies on combining (meshing) affordances, not on associating abstract properties. Thus, the indexical hypothesis is a type of mental model theory (e.g., Johnson-Laird, 1983) in that mental representations of language are representations of a situation (or the affordances of a situation) rather than a mental representation of the language itself. Third, the process of combination (mesh) relies on intrinsic constraints on coherent action rather than on formal, extrinsic, mathematical, or logical rules. We used these differences to formulate the experiments.

Experiments 1 and 2 demonstrate that what people judge as sensible depends on the mesh of affordances rather than on formal properties determined by LSA. Experiment 3 demonstrates that affordances are taken into account when interpreting linguistic innovations such as "The newsboy porched the newspaper" (Clark & Clark, 1979). We draw two conclusions from this work. The first is that high-dimensional theories such as LSA and HAL are inadequate accounts of human meaning because the symbols (high dimensional vectors) are not grounded. The second conclusion is that a more promising account of meaning is based on embodiment theory.

EXPERIMENT 1

Landauer and Dumais (1997) applied LSA to text comprehension and the measurement of coherence. First, they determined an LSA vector representation for a sentence as the average vector of the words contained in the sentence. Then, coherence between sentences was measured as the cosine of the angle between the vectors. The coherence of a paragraph is the average cosine between successive sentences. This LSA measure of coherence is highly correlated with "empirical comprehension scores" (Landauer & Dumais, 1997). Landauer and Dumais summarize this finding by noting that

TABLE 1

Two Example Scenarios and LSA Values for Experiment 1

	LSA cosines	
	Sentence to setting	Central to distinguishing
Setting: Marissa forgot to bring her pillow on her camping trip.		
Afforded: As a substitute for her *pillow,* she filled up an old sweater with **leaves.**	.58	.08
Nonafforded: As a substitute for her *pillow,* she filled up an old sweater with **water.**	.55	.06
Related: As a substitute for her *pillow,* she filled up an old sweater with **clothes.**	.63	.24
Setting: Mike was freezing while walking up State Street into a brisk wind. He knew that he had to get his face covered pretty soon or he would get frostbite. Unfortunately, he didn't have enough money to buy a scarf.		
Afforded: Being clever, he walked into a store and bought a **newspaper** to cover his *face.*	.38	.06
Nonafforded: Being clever, he walked into a store and bought a **matchbook** to cover his *face.*	.42	.03
Related: Being clever, he walked into a store and bought a **ski-mask** to cover his *face.*	.41	.46

Note. Central concepts are italicized; distinguishing concepts are in boldface.

"LSA, by capturing the central meaning of the passages appears to reflect the differential relations among sentences that led to comprehension differences" (p. 231). Of course, this correlational result must be treated with some caution. It may well be the case that coherent passages tend to have adjacent sentences with words that have similar LSA vectors. Nonetheless, that does not strongly imply that the statistical factors that underlie LSA vector similarity (occurrence of words in similar contexts) are the same factors that underlie coherence among sentences. We decided to examine this correlational result in both an experimental context and a correlational context.

We constructed 18 scenarios; two examples are given in Table 1. Each scenario began with one or more context-setting sentences. There were three versions of a critical sentence. In the Afforded version, the affordances of the objects could be meshed to result in a coherent action that accomplished the character's goal. For example, Marissa could use her sweater stuffed with leaves as a pillow, and Mike could use the newspaper to protect his face from the wind. In the Nonafforded versions, the objects cannot be (easily) meshed to accomplish the character's goal. For example, it is difficult to envision how a matchbook can afford protection from a brisk wind. Importantly, these sentences were constructed to have similar LSA values. That is, the cosine of the LSA vector for the Afforded sentence compared to the context-setting sentences is virtually identical to the cosine of the LSA vector for the Nonafforded sentence compared to the context-setting sentences. Thus, to the extent that the cosines are a measure of coherence, and to the extent that LSA captures the central meaning of the passages, people ought to judge the sentences as equally sensible.

Each scenario also had a third critical sentence that we called Related. For this sentence, the affordances could be coherently meshed and there was a relatively large LSA cosine between two critical words within the sentence, the central concept and the distinguishing concept. Thus, "clothes" has a relatively large cosine

TABLE 2

Characteristics of the 18 Scenarios Used in Experiment 1 (Standard Deviations in Parentheses)

Critical sentence	LSA cosines: Sentence to setting	LSA cosines: Central to distinguishing	Syllables in distinguishing concept	Kucera and Francis Frequency of distinguishing concepts
Afforded	.42 (.20)	.06 (.06)	3.00 (1.33)	60.1 (107.6)
Nonafforded	.43 (.18)	.06 (.05)	3.06 (1.43)	45.4 (101.7)
Related	.43 (.17)	.25 (.12)	2.61 (1.24)	50.2 (69.4)

with "pillow" and "ski-mask" has a relatively large cosine with "face."

Method

Participants. The 24 participants were students enrolled in introductory psychology classes at the University of Wisconsin—Madison. They received extra credit in exchange for their participation.

Materials. A total of 18 scenarios were constructed.[3] For each scenario, several LSA cosine measures were determined.[4] The means and standard deviations of these measures are presented in Table 2. The first measure is the critical-sentence to setting-sentence cosine. The means for the three types of critical sentences were very similar. In fact, there was not a significant difference among the conditions, $F(2,34) = .38$, $MS_e = .001$. A cosine of .43 seems to be a fairly high degree of similarity (or coherence); the paragraphs examined by Landauer and Dumais (Fig. 5 in that article) appear to have average cosines of .18 to .25.

Because the cosines for a sentence are based on the average LSA vectors for the words in the sentence, and because the critical sentences differed by only a word or two, the cosines are constrained to be similar. Thus, we decided to analyze another LSA measure of relatedness or sensibility, namely the cosines between a central concept in the critical sentence and the concepts that distinguish among the critical sentences. The central concepts were chosen by consensus. Thus, for the *Mike* example in Table 1, the central concept for the critical sentence was face, and the distinguishing concepts were newspaper, matchbook, and ski-mask. The mean cosines between the LSA vector for the central concept and the LSA vectors for the distinguishing concepts are given in Table 2 in the central-to-distinguishing column. There are significant differences among these cosines, $F(2,34) = 40.82$, $MS_e = .005$. The difference between the mean for the Related condition compared to the average of the Afforded and the Nonafforded conditions was significant, $t(17) = 6.84$, $SE = .03$. However, the difference between the Afforded and the Nonafforded conditions was negligible, $t(17) = .05$, $SE = .01$.

Table 2 also presents data on the number of syllables in the distinguishing concepts and the Kucera and Francis (1967) frequencies of the distinguishing concepts (the sum of the frequencies of the singular and plural forms). For number of syllables, $F(2,34) = 1.69$, $MS_e = .62$. For frequency, $F(2,34) = .11$, $MS_e = 9084$.

Procedure. Each participant received two forms, one for making sensibility judgments and one for making envisioning judgments. The order in which the forms were used was counterbalanced over the participants. Each form presented the 18 scenarios; that is, the context-setting sentences as well as the three critical

[3] All materials may be found at http://psych.wisc.edu/glenberg/jml_g&r.html.

[4] The LSA website at http://lsa.colorado.edu/ was accessed to calculate these cosines. We used the tasaAll space, in which the LSA matrix is based on the occurrence of 92,409 unique terms in 37,651 contexts selected from "texts, novels, newspaper articles, and other information." The SVD decomposition of this space yields vectors with a maximum of 419 dimensions. The scenarios and LSA values are available at http://psych.wisc.edu/glenberg/jml_g&r.html.

TABLE 3

Sensibility and Envisioning Ratings for Experiment 1

Critical sentence	Sensibility	Envisioning
Sensibility ratings first		
Afforded	4.32	4.94
Nonafforded	1.31	1.58
Related	6.10	6.40
Envisioning ratings first		
Afforded	4.84	4.88
Nonafforded	1.19	1.47
Related	6.52	6.17

sentences. The order of the critical sentences was counterbalanced so that all six orders of the three sentences occurred equally often. Participants were instructed to read the context-setting sentences and then to judge each of the critical sentences within that context. The sensibility judgments used a scale of 1 (virtual nonsense) to 7 (completely sensible). The envisioning ratings used a scale of 1 (impossible to imagine) to 7 (easy to imagine).

Results

The means for the two types of ratings are presented in Table 3. The ratings were analyzed using a two-factor analysis of variance. *F* ratios reported as $F1$ are when using participants as the random variable, and *F* ratios reported as $F2$ are when using texts as the random variable. We used a Type I error rate of .05 for all analyses. The first factor, First Judgment (Sensibility rating first or Envisioning rating first), was manipulated between subjects, and the second factor, Condition (Afforded, Nonafforded, and Related), was manipulated within subjects. Consider the sensibility ratings first. There was a large effect of Condition, $F1(2,44) = 682.94$, $MS_e = .23$; $F2(2,34) = 160.37$, $MS_e = 1.48$. Importantly, although LSA values cannot distinguish between the Afforded and the Nonafforded conditions, people can, $t1(23) = 18.80$; $t2(17) = 11.30$. Also, there was a significant difference between the Afforded and Related conditions, $t1(23) = 12.50$; $t2(17) = 5.11$. Finally, there was also a marginally significant interaction between the two factors. After people had performed the envisioning ratings, the sensibility judgments for the Afforded and Related sentences increased, whereas the sensibility judgments for the Nonafforded sentences decreased, $F1(2,44) = 3.01$, $MS_e = .23$, $p = .06$; $F2(2,34) = 14.17$, $MS_e = .07$.

The same type of analyses were conducted for the Envisioning ratings. Again, there was a large effect of Condition, $F1(2,44) = 677.93$, $MS_e = .21$; $F2(2,34) = 158.97$, $MS_e = 1.36$. The difference between the Afforded and the Nonafforded conditions was quite reliable, $t1(23) = 25.96$; $t2(17) = 12.29$, as was the difference between the Afforded and the Related conditions, $t1(23) = 10.00$, $t2(17) = 4.54$. The interaction was not significant, $F1$, $F2 < 1$.

We also examined how well the LSA ratings could predict sensibility ratings within the categories of Afforded, Nonafforded, and Related sentences. To compute these correlations, we obtained the mean Sensibility and Envisioning ratings for each of the 18 Afforded, Nonafforded, and Related Sentences and correlated these ratings with the LSA central-to-distinguishing cosines. These correlations are presented in Table 4. None of the correlations with the LSA cosines was significant. The low correlations are not because there is no variability in the LSA ratings (see Table 2). In fact, the LSA cosines correlated fairly highly with each other. For example, the correlations between the Afforded central-to-distinguishing LSA cosines and the Nonafforded central-to-distinguishing cosines was .59, and the correlation between the

TABLE 4

Correlations with Sensibility Ratings for Experiment 1

	LSA sentence to setting	LSA central to distinguishing	Envisioning ratings
Afforded	.04	.07	.96*
Nonafforded	.18	−.15	.90*
Related	−.24	.31	.94*

* $p < .05$.

Afforded and Related cosines was .50. In contrast to the failure of LSA cosines to predict the Sensibility ratings, the correlations between the Envisioning ratings and the Sensibility ratings were highly significant, ranging from .90 to .96.

Discussion

People can distinguish between sensible (Afforded) and less sensible (Nonafforded) sentences quite easily, even when LSA analyses suggest that the sentences are equivalent. Furthermore, the LSA measures do not strongly (or significantly) predict the sensibility ratings, whereas envisioning ratings do. These data support the Indexical Hypotheses and the idea of embodied meaning. Apparently, sentences make sense when the affordances can be meshed as directed by the syntax of the sentence. Nonetheless, the Afforded sentences were rated as less sensible than the Related sentences. We think that there is a simple explanation for this difference. According to the Indexical Hypothesis, meshing depends on an individual's experiences. Thus, if a person has experienced newspapers ripping apart in a stiff wind, that person will have a difficult time understanding how Mike could successfully use a newspaper to block the wind. Or, if someone sleeps so lightly that his or her rest would be disturbed by the crackling of leaves, then that person would have a difficult time understanding how Marissa could successfully use a leaf-stuffed sweater as a pillow. Given these differences in personal experiences, habits, and bodies, it is not surprising that not everyone would derive the same understanding of the sentences that the experimenters did when constructing the stimuli.

An alternative interpretation of the data is that we simply tapped into world knowledge: That is, people have learned that leaves can be used for pillows and newspapers for scarves. Certainly, the fact that these objects afford particular actions is uncontroversial, but did the knowledge of this come about from experience in using newspapers as scarves? There are three reasons to believe that the answer is "no." First, when we created the scenarios, we tried to be creative: to use objects in ways that we had never or rarely experienced them used. Second, we think it extremely unlikely that if given a central concept (e.g., pillow or face) as a cue in a free-association test, that anyone would produce the distinguishing concepts (e.g., leaves or newspaper). Third, the LSA data make exactly this point. That is, low- (close to zero) LSA cosines indicate that two words appear in orthogonal contexts. Thus, given that the mean central-to-distinguishing cosine for the Afforded sentences is only .06, we can be certain that the central concepts and the distinguishing concepts have appeared in nearly orthogonal contexts. Nonetheless, the participants rated the Afforded sentences as very sensible.

Might these results have been obtained because the experimental participants are engaged in a type of explicit problem solving ("How can a newspaper stop the wind?") instead of reading? Our intuitive sense is that the Afforded sentence are read and understood just about as easily as the Related sentences and that they do not engender explicit problem solving. Furthermore, an experiment demonstrating that participants read the Afforded sentences as quickly as Related sentences is reported at http://psych.wisc.edu/glenberg/jml_g&r.html.

For most theories of meaning, the results are impossible or at least surprising: How can two concepts such as face and newspaper be meaningfully related (as is apparent in the ratings) if they are not semantically related (e.g., members of the same category) and if they are not associated by virtue of common experience? The embodiment framework provides an answer that is so intuitively obvious that it leads to F values over 600. The meaning of a word is not given by its relations to other words and other abstract symbols. Instead, the meaning of words in sentences is emergent: Meaning emerges from the mesh of affordances, learning history, and goals. Thus the meaning of the word "chair" is not fixed: A chair can be used to sit on, or as a step stool, or as a weapon. Depending on our learning histories, it might also be useful in a balancing act or to protect us from lions in a circus ring. A newspaper can be read, but it can also serve as a scarf. And, when rolled up, a newspaper can be used to reach under a

TABLE 5

Example Stimulus Set for Experiment 2 (LSA Sentence-to-Setting Cosines Are in Parentheses)

Setting: Kate was cleaning her kitchen on Sunday morning after a big party she had the night before.

Object 1: Ceiling tile
- Afforded: Since she couldn't reach the ceiling, she stuck her broom up in the air to try to get a piece of gum off her ceiling tile. (.35)
- Nonafforded: She got down on her hands and knees to scrape the beer stains off the ceiling tile. (.31)

Object 2: Floor tile
- Afforded: She got down on her hands and knees to scrape the beer stains off the floor tile. (.28)
- Nonafforded: Since she couldn't reach the ceiling, she stuck her broom up in the air to try to get a piece of gum off her floor tile. (.33)

bed to retrieve an errant slipper. One need not have previously read about newspapers retrieving slippers to know that it is sensible. Instead, knowledge of the affordances of newspapers obtained either from perception or from analogical perceptual symbols is critical. Then, one can determine if the affordances (e.g., of rolled up newspapers and errant slippers) can be meshed to accomplish the goal.

EXPERIMENT 2

In this experiment the participants made judgments of sensibility and envisionability of sentences within a context, much as in Experiment 1. Experiment 2 differed from Experiment 1 in several important ways, however. First, in Experiment 1, a given distinguishing concept was always afforded (e.g., newspaper) or nonafforded (e.g., matchbook). Thus, one might argue that we just happened to pick nonafforded objects that were in some way less sensible than the afforded objects. In Experiment 2, a particular object was equally often afforded and nonafforded. Also, we constructed sets of objects that were, even when nonafforded, similar to the afforded objects. The objects in a set were differentiated only by a modifier. For example (see Table 5), the two objects in one set were "ceiling tile" and "floor tile." Both of these objects were used equally often as afforded and nonafforded objects. Second, in Experiment 1, for each scenario a participant judged the Afforded, Nonafforded, and Related sentences successively (albeit in a counterbalanced order). The contrast between the sentences may have induced a task demand to use different values of the rating scales. In Experiment 2, for each scenario a participant rated only one sentence relative to the context sentences. Third, we used a new set of stimuli to obtain different LSA ratings so that our results would not be dependent on just a few, perhaps unrepresentative, sentences.

Method

Participants. The 40 participants were students enrolled in introductory psychology classes at the University of Wisconsin—Madison. They received extra credit in exchange for their participation.

Materials. A total of 17 scenarios were constructed (see Table 5 for an example; all scenarios are available at http://psych.wisc.edu/glenberg/jml_g&r.html). Each scenario consisted of a context-setting sentence followed by one of four critical sentences. Two of the critical sentences used Object one (e.g., ceiling tile), and two critical sentences used Object two (e.g., floor tile). Object one and Object two were described by phrases using the same head noun but different modifiers. Each object was used in an Afforded sentence once and a Nonafforded sentence once. Because of the various constraints on stimulus construction, some of our Nonafforded sentences can be envisioned with a bit of elaboration of the situation. For example, if one of Kate's ceiling tiles had been removed from the ceiling, then one could easily imagine how Kate could get on her hands and knees to scrape off the beer stains. Thus,

TABLE 6

Sensibility and Envisioning Ratings for Experiment 2

Critical sentence	Sensibility	Envisioning
Sensibility ratings first		
Afforded	5.34	6.12
Nonafforded	2.25	2.88
Envisioning ratings first		
Afforded	5.89	5.88
Nonafforded	1.96	2.92

Afforded and Nonafforded should be treated as relative rather than absolute.

For each scenario, we determined the LSA cosine between the context-setting sentence and each of the four critical sentences. The mean for the Afforded sentences was .31, and the mean for the Nonafforded sentences was .30. This difference was not significant, $F(1,16) = 1.18$, $MS_e = .001$. There were no significant differences due to the two objects or the interaction of object and affordance condition, both Fs < 1.

The scenarios were presented in an arbitrary, but fixed, order. On one of the four stimulus forms, approximately four of the scenarios included Object one in the Afforded condition, four included Object one in the Nonafforded condition, four included Object two in the Afforded condition, and four included Object two in the Nonafforded condition. We then created three additional stimulus forms so that across the four forms a given scenario appeared equally often with each critical sentence. The four stimulus forms were used for both Sensibility ratings and for Envisioning ratings by simply changing the instructions printed on the top of the form (using the same scales as in Experiment 1). Combinations of stimulus forms generated eight between-subject groups formed by the factorial combination of form (1–4) and order of the ratings (Sensibility ratings before the Envisioning ratings or the reverse order).

Procedure. After signing consent forms, participants were randomly given one of the eight sets of forms. Instructions for the ratings were essentially identical to those used in Experiment 1.

Results

The means for the two types of ratings are presented in Table 6. The ratings were analyzed using a three-factor analysis of variance. The first factor, First Judgment (Sensibility rating first or Envisioning rating first) was manipulated between subjects. The within-subject factors were Condition (Afforded, Nonafforded) and Object (One or Two). Consider the Sensibility ratings first. There was a large effect of Condition, $F1(1,38) = 175.59$, $MS_e = 2.81$; $F2(1,16) = 54.16$, $MS_e = 7.82$. No other effects were significant. Once again, although the LSA values were virtually identical, people had no trouble discriminating between the Afforded and Nonafforded sentences.

The same types of analyses were conducted for the Envisioning ratings. Again, there was a large effect of Condition, $F1(1,38) = 294.51$, $MS_e = 1.30$; $F2(1,16) = 80.44$, $MS_e = 4.10$. No other effects were significant in analyses by both participant and text.

We also examined the relation between the LSA cosines and the participant's Sensibility ratings. The correlations were based on the 34 LSA sentence-to-setting cosines for the Afforded sentences (i.e., two afforded sentences for each of the 17 settings), the 34 LSA sentence-to-setting cosines for the NonAfforded sentences, the average Envisioning ratings for the sentences, and the average Sensibility ratings for the sentences. The results are in Table 7. The LSA cosines for the Afforded sentences did not correlate with the Sensibility ratings of the Afforded sentences. However, the LSA ratings for the Nonafforded sentences did significantly correlate with the Sensibility ratings for the Nonafforded sentences. As we reported for

TABLE 7

Correlations with Sensibility Ratings for Experiment 2

	LSA sentence to context	Envisioning ratings
Afforded	−.08	.67*
Nonafforded	.49*	.88*

* $p < .05$.

Experiment 1, the Envisioning ratings were highly correlated with the Sensibility ratings for both the Afforded and the Nonafforded sentences.

Discussion

The results are very similar to what we observed in Experiment 1. Namely, people can easily and reliably discriminate between the sensibility of the Afforded and the Nonafforded sentences even though the average LSA cosines (ostensibly a measure of meaning and coherence) are virtually identical for the two conditions. Furthermore, the Envisioning ratings strongly correlate with the Sensibility ratings, as would be expected if people judge sensibility by how well affordances can be meshed. Clearly, the results from Experiment 1 were not due to an odd choice of stimuli or task demands.

EXPERIMENT 3

This experiment presents a new type of challenge to high-dimensional theories of meaning as well as to many other formal theories of language comprehension. We examined people's understanding of denominal verbs; that is, verbs made out of nouns (Clark & Clark, 1979). These verbs are illustrated by the sentences "John bicycled to town" (from Clark & Clark, 1979) and "Ray toilet papered the front yard." As Clark and Clark note, these sorts of verbs are extremely common in English; Clark and Clark present a classification of more than 1300 denominal verbs. More interesting for our purposes are what Clark and Clark call innovations: denominal verbs made up (and understood) on the spot. Two examples from Clark and Clark are "The newsboy porched the newspaper" and "My sister Houdini'ed her way out of the locked closet."

Clark and Clark propose an innovative denominal verb convention that, supposedly, is used in understanding these innovative verbs. The convention is that when using such a verb, a speaker means to denote (a) the kind of situation that (b) the speaker believes that the listener can, on this occasion, (c) easily and (d) uniquely compute on the basis of (e) mutual knowledge so that (f) the parent noun plays one role in the situation and the remaining surface arguments of the denominal verb denote other roles in the situation. Of particular interest is the format of the mutual knowledge and how the computations are achieved. Sometimes it is sufficient to simply treat the noun in a standard sense. Thus, "to bottle beer" means to use bottles in their ordinary capacity. A semantic memory listing of the ordinary uses of a bottle might allow one to understand this use of the denominal verb. However, Clark and Clark note that understanding many innovations requires knowledge of unique characteristics of individuals, time, place, and physical properties. In other words, something very close to affordances. Thus, affordances of the objects need to be derived from physical properties to interpret "bottled" in, "We were stoned and bottled by the spectators as we marched down the street" (from a BBC broadcast as quoted in Clark & Clark).

Now, how is the computation [part (d) of the convention] performed to reach an understanding? As Clark and Clark note, salient characteristics of the noun and the situation are critical, and we must be able to combine the parent noun in one role with roles given by other parts of the sentence. But, it is not sufficient to just put the various parts of the sentence into various roles. We think that the comprehender must be able to consider the affordances of the various objects and the (action-based) goals that need to be accomplished in order to form a coherent (meshed) idea. Consider, for example, "She booked the leg" (see Table 8). In the first (Afforded) context, the goal is to level a table with a short leg. The affordances of a book allow it to be used to accomplish this goal with a coherent set of actions: put the book under the short leg of the table. In the second (Nonafforded) context, the goal is to find a book. In this context, it is not clear how the affordances of a book can be meshed with the actions of looking to generate a coherent set of actions for "She booked the leg."

Understanding sentences such as "She booked the leg" should be a challenge for high-dimensional theories because interpretation seems to be so dependent on specifics of context

TABLE 8

Example Stimuli for Experiment 3 (LSA Sentence-to-Setting Cosines Are in Parentheses)

Conventional verb (slimed), Afforded

Kenny sat in the tree house and patiently waited. He clutched the jar of green ooze in his hand, and watched the approaching school bus move closer to his house. The teenage girl stepped off and walked towards the tree house unaware of the little boy above her taking the cap off the jar. Kenny waited until she was directly beneath him, and an evil grin spread across his face. Then, Kenny slimed his sister. (.21)

Semi-innovative verb (booked), Afforded

Lori loved her new table, until she noticed that everything she placed on it slid off to the left. The left back leg was lower than all the others. She could not imagine how to fix the slant. Then she spotted a pile of hard covered books in the corner. She booked the leg. (.61)

Semi-innovative verb (booked), Nonafforded

Lori was having a really bad day. She could not find her textbook and she was late for class. Frantically, she ran over to the table where there was a pile of books. On the way, she banged her leg on the chair. She booked the leg. (.62)

Innovative verb (magazined), Afforded

Sebastian was perusing the latest issue of *Newsweek* when he was disturbed by a most annoying buzzing noise. He looked around the room to determine the source of this disturbance, and saw that a fly was patrolling the vicinity. It's incessant buzzing was making Sebastian insane. He had no choice but to terminate with extreme prejudice. So, he rolled up his *Newsweek* and waited patiently. When the fly came to rest on the coffee table in front of Sebastian, he recognized his opportunity. He magazined it. (.45)

Innovative verb (magazined), Nonafforded

Sebastian was perusing the latest issue of *Newsweek*. He became disturbed as he read an article about rising rates of home invasions in his vicinity. Sebastian decided to follow the advice of a security expert quoted in the magazine by purchasing a home security alarm. The salesman at the electronics store thought Sebastian was insane when he insisted on having the alarm installed that very day, but agreed when Sebastian threatened to terminate the sale. The alarm woke Sebastian when it began buzzing one evening.
He recognized his opportunity.
He magazined it. (.42)

and knowledge of physical properties. On the other hand, some aspects of the theories might be perfectly situated to handle these sorts of sentences. For example, one of Landauer and Dumais's (1997) major claims concerns how LSA can extract meaning from the flow of words. In particular, learning about words A, B, and C may affect knowledge of word D more than the simple presentation of D. For instance if the system encountered a novel unknown word which happened to have very similar patterns of contextual usage to ROAD and STREET's patterns, then the system's vectors for the new word would be similar to the vectors for ROAD and STREET. The innovative denominal verbs are a rough analog: We know about books from many different contexts. Based on those many occurrences, can LSA extract a new meaning? Or to put it a bit differently, is the meaning of "booked" implicit in the use of "book" in many contexts?

In the experiment we had people read a context and then a critical sentence containing a denominal verb. The participants judged the sensibility of the sentence and wrote a paraphrase of the sentence.[5] We varied the type of denominal verb as well as the type of context, although we did not use a full factorial design. One-third of the sentences used a Conventional denominal verb, such as "drummed." These Conventional denominal verbs always appeared in a context that affords the actions. Because the context supports the usual interpretation, and

[5] Participants also judged the degree to which the sentences were grammatical. The data did not seem to differentiate the theories and so this variable is not discussed further. The data are available at http://psych.wisc.edu/glenberg/jml_g&r.html, however.

because the verb is, presumably, part of the language, this condition forms a baseline. Another third of the critical sentences contained an Innovative denominal verb; that is, one that is not a standard part of English. The remaining sentences contained a Semi-innovative denominal verb. These verbs, such as "booked" in "She booked the leg," have a conventional denominal sense (e.g., to make a reservation), but we used the verb to convey a new sense (e.g., to balance a table by putting a book under one leg). Thus, our innovative use of "to book" is analogous to the innovative use of "to bottle" in the quote from the BBC.

The Innovative and Semi-innovative denominal verbs were factorially combined with two types of contexts, Afforded and Nonafforded. The Afforded contexts were written to suggest a goal that could be accomplished by meshing the affordances of the object named by the denominal verb (e.g., a book) with affordances of actions and other objects (e.g., a short table leg). The Nonafforded contexts were written to include many of the same words as the Afforded context, but to suggest a goal that could not be accomplished by meshing the affordances of the object named by the denominal verb. The Afforded and Nonafforded contexts were written so that the LSA cosines between these contexts and the critical sentence were approximately equated.

Predictions from LSA are straightforward. Because the Afforded and Nonafforded contexts are equally related to the critical sentences, the critical sentences should be seen as equally sensible and understood with equal ease. Also, within the Afforded and Nonafforded contexts, the greater the LSA cosine, the greater the judged sensibility ought to be. We were uncertain whether LSA would be more likely to make successful predictions about the Semi-innovative verbs (that do appear as verbs or modifiers in the language, but with a meaning that may interfere with the meaning required in the experiment) or the Innovative verbs (that do not appear as verbs in the language, so there is no interference).

Predictions from the Indexical Hypothesis are different. First, people should judge the sentences in the Afforded contexts as more sensible than those same sentences in the Nonafforded contexts. Furthermore, people should be able to paraphrase the sentences more effectively in the Afforded contexts. That is, in the Afforded context a critical sentence ought to be meaningful because it suggests a coherent set of actions; that is, a coherent idea. Thus, people should be able to paraphrase that idea. In the Nonafforded contexts, a critical sentence may appear bizarre and meaningless; that is, people will not be able to envision a coherent set of actions that underlie the sentence. In this case, people should have a difficult time producing a paraphrase.

Method

Participants. The 42 participants were students enrolled in introductory psychology classes at the University of Wisconsin—Madison. They received extra credit in exchange for their participation.

Materials. We wrote 18 critical sentences (all sentences and contexts are available at http://psych.wisc.edu/glenberg/jml_g&r.html). Six of these used Conventional denominal verbs, six used Semi-innovative denominal verbs, and six used Innovative denominal verbs. The Conventional denominal verbs met either or both of two criteria. One criterion was that the verb appeared in standard dictionaries as a verb and the dictionary meaning of the verb was the same as the sense used in the critical sentence. The second criterion was that the verb was part of the standard undergraduate lexicon in the unanimous judgment of the four people constructing the materials. Thus, "slimed" in "Kenny slimed his sister" was accepted as a Conventional denominal verb. The Semi-innovative denominal verbs appeared in standard dictionaries, but the dictionary meaning of the verb was different from that intended in the critical sentence. The Innovative denominal verbs met two criteria. One criterion was that the verb did not appear in standard dictionaries. The second criterion was that the verb was not in common usage in the undergraduate population.

For each of the Conventional verb sentences we wrote a multisentence context. The context was written so that the critical sentence appeared to follow naturally. That is, from the point of view of embodiment theory, the critical sentence was afforded. The mean LSA cosine between the critical Conventional sentences and their contexts was .33 (with a standard deviation of .19). For each of the Semi-innovative and Innovative critical sentences we wrote an Afforded context and a Nonafforded context as previously described. For the Semi-innovative sentences, the mean LSA cosine with the Afforded context was .47 (.21), and the mean with the Nonafforded context was .46 (.22). For the Innovative sentences, the mean LSA cosine with the Afforded context was .45 (.11), and the mean with the Nonafforded context was .47 (.11). There were no significant differences in an analysis of these data (all Fs $<$ 1).

Two forms were constructed. On the first form, the scenarios were typed in an arbitrary order except that successive groups of three included one exemplar of each type of verb. Also, successive pairs of Semi-innovative verb sentences included one in an Afforded context and one in a Nonafforded context. The same was true for successive pairs of Innovative verbs. We counterbalanced the occurrence of Afforded and Nonafforded contexts across the two forms. Thus, a Semi-innovative or Innovative verb that appeared in an Afforded context on one form appeared in the Nonafforded context on the other form. The Conventional verbs always appeared in an Afforded context. The two forms were randomly assigned to participants so that 21 participants received each form.

The participants were instructed,

> In this experiment we are studying how people understand sentences with unusual and unfamiliar words. Here is what we would like you to do for each of the 18 paragraphs written below. 1) Read a paragraph to get a good idea of what it is about. 2) Judge if the last sentence is grammatical. That is, does the sentence have the right number of nouns and verbs in the appropriate order, and do the parts of the sentence have the appropriate relations, such as singular subject with a singular verb. To make this judgment, use the grammaticality scale from 1 (not grammatical) to 7 (completely grammatical). 3) Judge if the last sentence makes sense in the context of the paragraph. To make this judgment, use the sensibility scale from 1 (virtual nonsense) to 7 (completely sensible). 4) Write a paraphrase of the last sentence. That is, use different words to describe what the last sentence states.

TABLE 9

Sensibility Ratings and Paraphrase Scores for Experiment 3 (Standard Errors Are in Parentheses)

	Sensibility rating	Paraphrase
Conventional verbs		
Afforded	5.67 (.12)	.99 (.01)
Semi-innovative verbs		
Afforded	3.78 (.27)	.96 (.02)
Nonafforded	2.29 (.21)	.13 (.03)
Innovative verbs		
Afforded	4.12 (.24)	.96 (.02)
Nonafforded	2.06 (.16)	.32 (.03)

Results

The data of main interest are in Table 9. Consider first the Sensibility ratings for the Semi-innovative and Innovative verbs. Contrary to the expectation derived from the LSA cosines, the type of context (Afforded versus Nonafforded) had a large effect on the ratings, $F1(1,41) = 98.43$, $MS_e = 1.34$; $F2(1,5) = 54.03$, $MS_e = .35$. There was also a significant interaction between verb type and context, $F1(1,41) = 7.82$, $MS_e = .42$, but it was not significant in the analysis by text, $F2(1,5) = 2.18$, $MS_e = .22$. As predicted by the Indexical Hypothesis, the Innovative verbs were judged sensible in the Afforded condition (4.12) and pretty much nonsense in the Nonafforded condition (2.06). Compared to the Innovative verbs, the Semi-innovative verbs were judged a bit less sensible in the Afforded condition (3.78), and a bit more sensible in the Nonafforded condition (2.29), in which the standard meaning of the verb might be useful. Although the Innovative verb in the Afforded condition was judged as more sensible than in the Nonafforded condition, it was not judged as sensible as the Conventional verb in the Afforded condition, $t1(41) = 9.25$; $t2(10) = 3.68$.

Consider next the paraphrase data. We scored the paraphrase data simply. If a paraphrase had the same meaning as the intended meaning of the sentence in the Afforded condition, it was scored as a 1. If the paraphrase had a different or undetermined meaning, it was scored as a zero. The analysis of variance demonstrated a main effect for type of verb in the analysis by participants, $F1(1,41) = 16.01$, $MS_e = .03$, but not in the analysis by text, $F2(1,5) = 1.27$, $MS_e = .05$. Contrary to the predictions from LSA theory, there was a main effect for type of context, $F1(1,41) = 666.62$, $MS_e = .04$; $F2(1,5) = 143.67$, $MS_e = .02$. When the sentences were in the Afforded context, people overwhelmingly found them to mean the same thing that we intended them to mean. However, those same sentences in the Nonafforded condition were paraphrased differently. There was also an interaction between the two factors in the analysis by participants, $F1(1,41) = 12.78$, $MS_e = .03$, although not in the analysis by text, $F2(1,5) = 1.58$, $MS_e = .03$. For the Innovative verbs in the Afforded condition, the paraphrases were almost always consistent with what we expected, whereas in the Nonafforded condition, people wrote paraphrases consistent with the afforded meaning 33% of the time. For the Semi-innovative verbs in the Afforded condition, the paraphrases were again highly consistent with what we expected, whereas in the Nonafforded condition, people wrote paraphrases consistent with the Afforded meaning only 13% of the time. The difference between 33 and 13% arises because the participants sometimes used the alternative meaning of the Semi-conventional verb in the Nonafforded condition, so that the paraphrase would not match the meaning of the verb in the Afforded condition. Note that the accuracy of paraphrasing the Innovative verb in the Afforded condition (.96) is about the same as the accuracy of paraphrasing the conventional verb in the Afforded condition (.99), $t1(41) = 1.43$; $t2(10) = .71$. Thus, participants were able to paraphrase the Innovative verbs in the Afforded condition very accurately, although some differences between these two conditions may be obscured by ceiling effects.

We also examined the correlation between the LSA cosines and judged sensibility. Because there were only six verbs of each type (Conventional, Semi-innovative, and Innovative), we do not report the correlations separately for each type of verb. For the 18 Afforded sentences (six of each type), the correlation was significantly negative, $r = -.55$. For the 12 Nonafforded sentences (six Semi-innovative and six Innovative) the correlation was not quite significant, $r = -.47$. Of course, the negative correlations are just the reverse of what the LSA theory predicts.

Discussion

People can understand innovative denominal verbs when the affordances of the named object can be meshed with affordances of other objects to accomplish goals. The sensibility of these sentences is rated as high, and participants' paraphrases of the afforded sentences are highly accurate. In many ways this is not surprising. As Clark and Clark (1979) noted, "Forming and understanding [innovations] is . . . an intrinsic part of our capacity to use language, and should be accounted for by any theory of language that claims to be complete" (p. 809). Unlike people, however, the LSA cosines do not discriminate between Afforded and Nonafforded conditions. More vexing for the theory is that for just those innovations that are understood (those in the Afforded condition), the LSA cosines are negatively, rather than positively, correlated with sensibility judgments.

GENERAL DISCUSSION

The data from the three experiments speak strongly against the claim that high-dimensional vector representations derived from the language stream can be an adequate account of human meaning. People can consistently (Experiments 1 and 2) discriminate between sentences that describe afforded actions and those that attempt to describe nonafforded actions (e.g., using glasses to dry one's feet); LSA cannot. People can understand innovations when they describe meshed affordances (Experiment 3); LSA cannot.

There are at least four arguments that could be offered in defense of LSA as a theory of

TABLE 10

Effects of Changing Number of Dimension on the LSA Sentence to Context Cosines

	Number of dimensions			
	50	100	200	419
Marissa text (see Table 1)				
Afforded	.91	.85	.76	.58
Nonafforded	.85	.76	.68	.55
Related	.96	.91	.82	.63
Mike text (see Table 1)				
Afforded	.73	.64	.53	.38
Nonafforded	.75	.66	.57	.42
Related	.74	.66	.56	.41

human meaning. The first argument is based on the number of dimensions in the vectors representing each word. Landauer and Dumais (1997) demonstrated (their Fig. 3) an inverted U-shaped relation between number of dimensions and proportion correct choices on a synonym test. We did not vary the number of dimensions; instead, we used the default value of 419 dimensions. Judging from Landauer and Dumais' Fig. 3, 419 dimensions is well within the range of dimensions that support strong performance on the synonym test. To examine the issue a bit more closely, Table 10 presents illustrative data regarding the effects of changing number of dimensions. The two examples are taken from Table 1. Whereas changing the number of dimensions produced large changes in the LSA cosines, the relative orderings stayed constant. For the Marissa sentence, the Afforded sentence cosine was always a bit larger than the Nonafforded sentence cosine. For the Mike sentence, just the opposite pattern occurred. Because over all of our stimuli the differences between the Afforded and Nonafforded LSA values (with 419 dimensions) were very similar, and because across the stimuli sometimes the Afforded cosines were the larger and sometimes the Nonafforded were the larger, we think it unlikely that changing the dimensionality of the vectors would affect our conclusions.

The second argument is based on choice of semantic space. In LSA, the SVD procedure is performed on a matrix relating words to contexts, and selection of the appropriate context may be critical. The semantic space that we used, the tasaAll space, was the most general space (see footnote 3) of those available at the LSA website. The tasaALL space is based on a wide domain of material including both expository and narrative materials; roughly half of the paragraphs in the corpus come from the area of language arts. Nonetheless, that does not make it the most appropriate space. We offer four counters to the argument that our results depend on using an inappropriate space. First, it is hard to imagine that people maintain independent spaces for overlapping domains of knowledge. Second, when using a different semantic space, the LSA values would have to change dramatically to effect the results. It seems unlikely that any semantic space would produce a large difference between the Afforded and Nonafforded conditions given that the sentences differ by just a single word (in Experiments 1 and 2). Third, if the LSA values did change dramatically with the space, and the selection of space is not specified by the theory, the theory is rendered untestable. The fourth counter relies on the reader to participate in a version of Experiment 1. One of the spaces is the encyclopedia space which is based on 30473 encyclopedia articles, several of which touch upon the topic of classical conditioning. Within this space, consider the LSA cosines relating the following context-setting sentence and the two alternative sentences.

Context-setting: Classical conditioning is a procedure that can induce a type of learning.
Alternative 1: The procedure is based on pairing two types of stimuli.
Alternative 2: The procedure is based on pairing two types of Pavlov.

The cosine relating Alternative 1 and the context-setting sentence is .36 (based on 300 dimensions), whereas the cosine relating Alternative 2 and the context-setting sentence is .38. Regardless of the similarities of the cosines, we predict that readers will find Alternative 1 to be more sensible and more coherent within the context than Alternative 2.

The third argument in defense of LSA is that our experiments unfairly tested LSA by using words in novel ways (e.g., stuffing sweaters with leaves and magazining flies). Surely, so the argument goes, if LSA had had experience with these novel locutions, it would have done just fine. This argument requires three rejoinders. (a) In Experiments 1 and 2, words were used to describe novel scenarios, but the "meanings" of the words (e.g., "sweater" and "leaves") were completely ordinary. Also, although we chose the scenarios to be novel, they do not strike us as in any way bizarre. Instead, we used the language in a way that it is supposed to be used: to convey new information. (b) In regard to Experiment 3, LSA's predictions were disconfirmed for both the innovative and the semi-innovative denominal verbs. LSA has had experience with the semi-innovative denominal verbs (such as "to book") because the majority of them are standard in English. (c) A computational model should be able to account for material beyond its training set. It is especially important that a theory of language and meaning be flexible and productive beyond its training set because humans are flexible and productive. As a case in point, our participants had no experience with sweaters stuffed with leaves (the experimenters intentionally generated novel scenarios), but the participants had little difficulty understanding these sentences from outside their training sets.

The fourth argument in defense of LSA is that it is a theory of word meaning, not how word meanings combine. Indeed, Burgess has been clear that HAL is a theory of word meaning and that more is needed to turn it into a complete theory of a linguistic meaning. Broader claims have been made for LSA, however. As noted in the introduction, Landauer and Dumais (1997) model sentence coherence using LSA (see also, Foltz, Kintsch, & Landauer, 1998; and Landauer et al., 1997). In addition, consider this quote from Landauer (1999):

> Why has LSA accomplished as much as it has? I think the answer is straightforward; LSA (and HAL as well) has been able to objectively, and in large part correctly, represent and model the acquisition, representation, and combination of word meanings, and word meanings are of paramount importance in discourse. (p. 309)

There is also a more important point to be made: No matter how LSA, HAL, and other ungrounded symbol theories are extended and modified, ungrounded, arbitrary symbols cannot be an adequate basis for human meaning. The reason is that computational manipulation of abstract symbols merely produces more abstract symbols, not meaning. Searle's (1980) Chinese Room argument is one demonstration of this point. We now present a new argument that ungrounded symbols cannot, in principle, be combined in a manner that captures the meaning humans derive from combining words. Of course, our experiments demonstrated this empirically for LSA. The "in principle" argument is based on demonstrating the inadequacy of four procedures for attempting to use arbitrary symbols to discriminate between descriptions of novel combinations that people find sensible (e.g., covering a face with a newspaper to block the wind) and descriptions of novel combinations that people do not find sensible (e.g., covering a face with a matchbook). By a novel combination, we mean that the symbol system does not include a direct assertion relating the symbols that need to be combined, thus the assertion needs to be derived or verified. To focus the argument, consider asking a symbol system if the following novel combination is sensible: Can symbol 10011001 be put into relation 11110001 with symbol 10011010? First, consider attempting to use the symbols to do what people seem to do effortlessly, namely to imagine or simulate (Barsalou, 1999) the literal combination of shapes. Because the symbols are arbitrarily related to the objects, the symbols cannot literally be juxtaposed (e.g., "1111000110011010") to produce anything sensible. This is true even if the symbols included a very fine-grained coding of perceptual features because by definition those features are arbitrarily related to real shape.

Second, consider using arbitrary symbols that include a very fine-grained coding of perceptual features (e.g., a complete coding of the shape of a prototypical newspaper) that can be used to create a fine-grained perceptual image. Can this image be used to simulate the combination of the referents of the symbols? There are two reasons why this procedure cannot save arbitrary symbols. (a) If such an algorithm were developed, the real work of determining if the objects fit together is being done by the analogical perceptual representation, not the arbitrary symbols. (b) If such an image can be generated from the fine-grained coding of perceptual features, then, in fact, the coding of features is not arbitrary; instead the coding is related to perception by the function that generates the perceptual image. Hence, such a system does not use arbitrary symbols.

Third, consider the possibility of deriving through logical deduction that two objects fit together sensibly. Thus, to answer the question, "can symbol 10011001 be put into relation 11110001 with symbol 10011010" the symbol system brings to bear several thousands of facts about each of the two objects, such as "symbol 10011001 [one of the original symbols] can be put into relation 1110010 [a new relation] to symbol 11000000 [a new symbol]." Attempting to find a logical relation between the two original symbols results in a combinatoric explosion that would soon overwhelm even the most powerful computers. More devastatingly, given enough time and enough facts, a logical path can be found to connect almost any two objects. For example, a matchbook is made of cardboard. Cardboard can be recycled. Recycled cardboard can turned into a large sheet of newsprint . . . and so a matchbook can be used to cover the face. But, this is just the conclusion that most people fail to derive when reading sentences such as those used in Experiments 1 and 2. That is, when the symbol system is given enough facts, it derives conclusions that people do not.

Fourth, consider determining if two objects can combine using a parallel, constraint satisfaction algorithm (e.g., Kintsch, 1988) rather than logical deduction. The constraints are framed using arbitrary symbols, but now we can also add statistical information to facts, such as how frequently symbol 10011001 has a relation 1110010 to symbol 11000000. Unfortunately, there is no way to judge if the resulting (perhaps massive) set of linked constraints is sensible or not. That is, a constraint satisfaction system might be faster than logical deduction, but once enough linking constraints are found it would still end up noting that a matchbook can be used to cover the face, albeit with a low probability or low satisfaction of the constraints. Does the low-probability model a person's sense that matchbooks cannot be used to cover faces? Unfortunately, low or high constraint satisfaction cannot be used to judge sensibleness of novel combinations because the constraints may be contextually inappropriate. For example, breaking a pencil in half ruins it (a strong constraint against being able to use the pencil to do anything), and ramming that half pencil through a tennis ball ruins the ball (a strong constraint against being able to use the ball to do anything). Nonetheless, under duress a penciled ball makes a serviceable spinner for a board game.

A symbol (or vector) used by LSA or HAL is computed from a large number of strings of words. Does this amount to the vector being grounded in the strings of words? In some sense the answer is yes, but it cannot be used to save the theories from the arguments developed above. Because each word is (to the computer program) an ungrounded symbol arbitrarily related to its referent, piling on more and more relations cannot produce symbols that combine appropriately. To reiterate, the computational manipulation of abstract symbols merely produces more abstract symbols, not meaning.

In summary, abstract symbols arbitrarily related to their referents cannot, in principle, account for the way that human beings can effortlessly discriminate between sensible and nonsense novel combinations. Unfortunately, all formal theories in cognitive psychology are based on just those sorts of symbols. For just two examples, Masson (1995) presents a connectionist theory of semantic memory. In that theory, meaning is represented as a vector of 80 binary values, which are ungrounded symbols. In Shiffrin and Steyvers (1997) REM theory, "The lexical/semantic representation of a word consists of [20] non-zero feature values . . . ."

These values are ungrounded. The reason for using ungrounded symbols is clear: They are far easier to use in computer and mathematical simulations than are grounded representations. When symbols are ungrounded, however, none of the theories can discriminate between using a newspaper to cover one's face and using a matchbook.

Despite their inadequacy as theories of meaning, the HAL and LSA models have much they can contribute to the field of cognitive science because they can be used as tools for testing theories (see Perfetti, 1998, for an independent development of this argument). For example, Livesay and Burgess (1998) use the HAL model to test the Compound Cue theory (McKoon & Ratcliff, 1992) of mediated priming. The theory explains that priming is dependent on two factors, lexical co-occurrence and semantic relatedness. The HAL model is ideally situated to provide a metric of co-occurrence and relatedness, which previously could be only loosely specified. Similarly the HAL model can be used to demonstrate that proper names are distinct from common nouns on the basis of co-occurrence patterns in language (Burgess & Conley, 1998). This distinction is of importance to theoretical explanations of the privileged role proper names enjoy in language processing (Robertson, Gernsbacher, & Robertson, 1998). Similarly, we have used LSA as a tool to demonstrate that verbal background knowledge is unlikely to be able to discriminate between Afforded and Nonafforded sentences. In addition to providing useful tools for research, the HAL and LSA models may contribute to innovative applied tools. For example, AutoTutor (Graesser, 1998) is a computer-based tutoring system designed to meet the needs of individual learners. The system uses LSA calculations to automatically (albeit roughly) evaluate a particular student's understanding of a topic. The point is that these models have considerable value, even though they fail to account for how people understand language.

As discussed in the introduction, the Indexical Hypothesis moves toward solving the symbol grounding problem by taking a relatively simple step: It does not depend on abstract, arbitrary symbols whose meanings are based on relations to other symbols of the same sort. Instead, the hypothesis is that words are indexed to real objects or analog perceptual symbols. This is a step toward a solution of the symbol grounding problem in two ways. First, meshed representations are built out of the relation between action and the environment. Thus, what gives meaning to a situation is grounded in actions particularized for that situation. For example, the meaning of the cup on your desk is not given by relations among abstract symbols within a vast semantic network. Instead, the meaning of the cup is what you can do with it (drink out of it, throw it, or use it as a paperweight) given your current state and goals. Second, Barsalou (1999) discusses how representations constructed from perceptual symbols can be compared to perceptual experience. Because the perceptual symbols and the perceptual experiences are analogically related, we eliminate problems necessitated by trying to match to perceptual experience symbols that are arbitrarily related to that experience.

Nonetheless, it is abundantly clear that the Indexical Hypothesis requires greater specification of virtually all of its presumed processes. How is language parsed? How is a parsed phrase indexed to an object? (See Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995, for data indicating that people do index words to objects.) How are affordances derived? How does the mesh processes actually work? (See Kaschak & Glenberg, 2000, for how syntax guides mesh.) With all of these questions needing to be answered, why should the Indexical Hypothesis be seen as an alternative to computationally explicit and powerful abstract symbol theories? As we have demonstrated, ungrounded symbols cannot, in principle, be the basis of language comprehension. Hence, no matter how many experiments we conduct and no matter how many complications we add, in the end, they fail as theories of meaning. Embodied theories, and the Indexical Hypothesis in particular, may well be incorrect, but they are not doomed in principle.

In summary, we have argued (along with Barsalou, 1999; Harnad, 1990; Lakoff, 1987;

Searle, 1980; and many others) that the symbol grounding problem is pervasive in cognitive theories such as LSA and HAL. In fact, Edelman (1992) has concluded that the abstract symbol view of meaning "is one of the most remarkable misunderstandings in the history of science" (p. 228). In place of abstract symbol theory we have described the Indexical Hypothesis derived from an embodied theory of cognition (Glenberg, 1997; Glenberg & Robertson, 1999). This approach moves us nearer to solving the symbol grounding problem and because of that it gives us a way to understand how we know that shirts can be used to dry our feet, but glasses cannot.

REFERENCES

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences,* **22,** 577–660.

Barsalou, L. W., Solomon, K. O., & Wu, L.-L. (1999). Perceptual simulation in conceptual tasks. In M. K. Hiraga, C. Sinha, & W. Wilcox (Eds.), *Cultural, typological, and psychological perspectives in cognitive linguistics: The proceedings of the 4th conference of the International Cognitive Linguistics Association* (Vol. 3, pp. 209–228). Amsterdam: John Benjamins.

Burgess, C., & Conley, P. (1998). Developing semantic representations for proper nouns. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th annual meeting of the Cognitive Science Society* (pp. 185–190). Mahwah, NJ: Erlbaum.

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes,* **12,** 177–210.

Clark, E., & Clark, H. H. (1979). When nouns surface as verbs. *Language,* **55,** 767–811.

Edelman, G. M. (1992). *Bright air, brilliant fire.* New York: Penguin.

Ervin-Tripp, S. (1973). Some strategies for the first two years. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 261–286). New York: Academic Press.

Gibson, J. J. (1979). *The ecological approach to visual perception.* New York: Houghton Mifflin.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences,* **20,** 1–55.

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes,* **28,** 1–26.

Graesser, A. (1998, July). *AutoTutor: A computer tutor that simulates human tutorial dialogue.* Symposium presented at the eighth annual meeting of the Society for Text and Discourse, Madison, WI.

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

Horn, W. S. (1998). *Perceptual priming in simile comprehension.* Paper presented at the Meeting of the Midwestern Psychological Association, Chicago.

Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), *Foundations of cognitive science.* Cambridge, MA: MIT Press.

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language,* **43**(3), 508–529.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review,* **95,** 163–182.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown Univ. Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Landauer, T. K. (in press). Latent Semantic Analysis: A theory of the psychology of language and mind. *Discourse Processes,* **27,** 303–310.

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.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes,* **25,** 259–284.

Landauer, T. K., Laham, D., Rehder, B., & Schreiner, M. E. (1997). How well can passage meaning be derived without using word order: A comparison of Latent Semantic Analysis and humans. In (M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Meeting of the Cognitive Science Society* (pp. 412–417). Mahwah, NJ: Erlbaum.

Livesay, K., & Burgess, C. (1998). Mediated priming does not rely on weak semantic relatedness or local cooccurrence. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Meeting of the Cognitive Science Society* (pp. 609–614). Mahwah, NJ: Erlbaum.

Lund, K., Burgess, C., & Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. In *Proceedings of the Seventeenth Annual Meeting of the Cognitive Science Science Society* (pp. 660–665). Hillsdale, NJ: Erlbaum.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21,** 3–23.

MacWhinney, B. (1998). The emergence of language from embodiment. In B. MacWhinney (Ed.), *The emergence of language from embodiment.* Mahwah, NJ: Erlbaum.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **18,** 1155–1172.

Newton, N. (1996). *Foundations of understanding*. Philadelphia: John Benjamins.

Oden, G. C. (1984). Integration of fuzzy linguistic information in language comprehension. *Fuzzy Sets and Systems,* **14,** 29–41.

Oden, G. C. (1987). Concept, knowledge, and thought. *Annual Review of Psychology,* **38,** 203–227.

Perfetti, C. A. (1998). The limits of co-occurrence: Tools and theories in language research. *Discourse Processes,* **25,** 363–377.

Pinker, S. (1994). *The language instinct*. New York: William Morrow.

Robertson, D. A., Gernsbacher, M. A., & Robertson, R. R. W. (1998). Referential specificity, first-mention, suppression, enhancement, and memory for characters in sentences. Manuscript submitted for publication.

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

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review,* **4,** 145–166.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science,* **268,** 1632–1634.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard Univ. Press.

(Received March 17, 1999)
(Revision received November 3, 1999)