Journal of Memory and Language **43,** 402–408 (2000)
doi:10.1006/jmla.2000.2715, available online at http://www.idealibrary.com on IDEAL®

INVITED REPLY

Theory and Operational Definitions in Computational Memory Models: A Response to Glenberg and Robertson

Curt Burgess

University of California, Riverside

Glenberg and Robertson (2000) conducted three experiments to investigate the adequacy of high-dimensional memory models in accounting for the meaningful interpretation of sentences. They conclude that high-dimensional memory models (such as LSA and HAL) are inadequate as theories of meaning. As an alternative account, they offer an embodied approach to cognition—the indexical hypothesis. In this article, I detail how Glenberg and Robertson have failed to extend to the models they criticize the same considerations that are extended to the humans in their experiments. The model is evaluated without providing appropriate experience for the tests that are conducted. Moreover, testing a representational model (e.g., LSA) for processing considerations is inappropriate and ultimately contributes to what are serious flaws in their interpretation of their experimental results.

Semantics. The curse of man.
—Maxwell, *The Singer Enigma*

"Meaning is the most important problem in cognitive psychology. Meaning controls memory and perception. Meaning is the goal of communication" (from Glenberg & Robertson, 2000, p. 1). Glenberg and Robertson (hereafter GR) have succinctly characterized the centrality of meaning in our field and in their article they have presented a criticism of the high-dimensional memory (HDM) approach as inadequate as a theory of meaning representation. They present three experiments that they purport to offer as evidence that HDM models cannot represent meaning and offer an alternative approach, the indexical hypothesis, that avoids several important problems that are attributed to HDM models in general.

The reader of the GR article will find it superficially compelling—the experiments are methodologically solid (with one important exception), their examples are convincing, and the results are clear (perhaps to a fault). My goal in this article is to provide the reader with a discussion of how the arguments in GR are inadequately developed and how the conceptual linkage is not provided for how the indexical hypothesis should work. Finally, I discuss what is the cause of these difficulties—a lack of operational definitions of pivotal concepts in the indexical hypothesis. Throughout this article, I provide some clarification of the HDM approach and discuss how the HDM models can provide a clear answer to a number of the questions that GR raise. This said, however, it is important to realize that GR have presented a set of experiments that make clear some limitations of the HDM approach. They argue that these limitations are fatal with respect to how HDM can provide an account of meaning. I argue that these limitations are more likely

This research was supported by a NSF Presidential Faculty Fellow award (SBR-9453406) to Curt Burgess. I thank Chad Audet, Patrick Conley, Kevin Lund, Kay Livesay, and David Robertson for their helpful comments on this article. I also express my appreciation to Art Glenburg for his willingness and eagerness to talk for hours about these issues over the past 5 years. Web address for more information on the HAL model: http://hal.ucr.edu.

Address correspondence and reprint requests to Curt Burgess, Department of Psychology, Computational Cognition Lab, University of California, 1419 Life Science Bldg., Riverside, CA 92521. Fax: 909-787-3985. E-mail: curt@cassandra.ucr.edu.

0749-596X/00 $35.00

implementational shortcomings rather than theoretical flaws.

In all three experiments, the Internet-accessible version of the LSA model (http://lsa.colorado.edu; see Landauer & Dumais, 1997) was used to compute similarity scores for GR's stimuli. First, I briefly review the basic sentence-comprehension results offered by GR. Then the interpretation of these results is discussed in the context of how HDM models learn their representations.

GR'S EMPIRICAL RESULTS

Experiment 1

In this experiment, subjects read a sentence that provided a setting for an experimental sentence that followed. Sensibility and envisioning ratings were made by the subjects on three kinds of experimental sentences (an example is provided below).

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.
Nonafforded: As a substitute for her pillow, she filled up an old sweater with water
Related: As a substitute for her pillow, she filled up an old sweater with clothes

GR first analyzed the stimuli used in this experiment using the LSA website to compute cosines for the stimuli in the different conditions. First, they found the cosine between the vectors constructed from the setting sentence and the vectors constructed from the experimental sentences. Landauer and Dumais (1997) use these cosines to predict relatedness or coherence of sentences. GR found no differences in this measure of relatedness across the conditions. The second analysis examined cosines for each experimental sentence, where the cosine compared the vectors constructed from key words within each sentence. The cosines for the Afforded and Non-afforded sentences were not different from one another and close to zero. Those two conditions resulted in cosines lower than the cosines in the Related condition. A different pattern of results was obtained with the sensibility and envisioning ratings from the subjects on these sentences. GR obtained the rather predictable result—related and afforded sentences were judged as more sensible and envisionable than the nonafforded sentences. Obviously the related and afforded conditions made more sense in the context of the setting sentence—but GR showed that LSA was unable to detect this difference. As a result, GR conclude that LSA does not seem to capture meaning in a language setting, a task that is trivially straightforward for a person. GR argue that according to the indexical hypothesis "meshing depends on an individual's experiences."[1]

Experiment 2

This experiment is very similar to Experiment 1. The value of Experiment 2 is that the afforded or nonafforded objects are held constant. The basic pattern of results, and GR's conclusions, are similar to that of Experiment 1.

Experiment 3

This experiment involves the innovative use of denominalized verbs such as *booked* that can take on a new meaning "on the fly." The most relevant task the subjects were asked to do was to make sensibility ratings of the critical sentence after reading a context paragraph. The authors find that the afforded condition with the innovative verbs are rated as more sensible than the nonafforded conditions. Once again, GR find that LSA seems insensitive to these contextually novel uses of the denominalized verbs, whereas humans are able to figure this out.

GETTING MEANING INTO AN HDM MODEL

How HDM models, such as LSA or HAL, develop meaning representations involves contextual learning from language experience. The details are beyond the scope of this article and can be found elsewhere (Burgess & Lund, 2000; Landauer & Dumais, 1997). In characterizing the conceptual manipulation and language comprehension of humans, the nature of the expe-

[1] Somewhat curiously, the afforded condition is rated a little less sensible and envisionable than the related condition despite all the affording and meshing that is taking place.

rience and creativity in language usage are important issues. The basic flaw of the experiments conducted by GR lies in their failure to adequately extend two straightforward premises about human cognition to the HDM models that they attempt to investigate.

Premise 1: The nature of the experience that a person or a model has is a powerful predictor of performance. (In HDM models such as LSA or HAL, the language experience is what determines the nature of the memory representations.)

Premise 2: A person has an active and creative cognitive processing system to facilitate the use of the contents of memory.

One might expect proponents of an embodiment theory of cognition to be sensitive to how experience facilitates performance. GR state "meshing depends on an individual's experience." Accepting this assertion highlights a major flaw in the reasoning behind this entire set of experiments. The nature of the experience, i.e., the corpus of text, simply does not provide the LSA model with the necessary information to plausibly test the contexts and sentences that GR use. The text used by the LSA model implementation used by GR has no experience with most, if not all, of the situations presented by GR. What does it mean to say that LSA does not have the requisite information? Meaning in a high-dimensional model, whether of a word or of a larger text unit, hinges on the contexts in which words occur. GR used an implementation of LSA that was trained on the TASA corpus which lacks the experiences (contexts) that GR's stimuli test. For example, there is no experience of filling up sweaters with clothes for a pillow or covering one's face with a newspaper (see GR's stimuli). Given the absence of these kinds of contextual learning histories, LSA will not fare well with the stimuli in these experiments. GR argue that this absence is no excuse for the failure of LSA since LSA did have experience with the particular words they used. However, to reiterate, it is the contextual experience that is important for learning language, not simply the exposure to words. LSA did not have a contextual history with the novel word usage included in this set of experiments. However, as GR point out, neither did their subjects who were able to understand the afforded contexts. This leads us to the next critical issue with these experiments.

HUMANS ARE SMARTER THAN LSA OR HAL!

There is another important issue, related to Premise 2. GR note that their subjects likely did not have the exact experiences described in these experimental situations; however, they rated the afforded and related conditions as more sensible. HDM models do not have an active online creative processing component. The typical implementation of LSA and HAL for testing is that of a representational model. One can use them to compute a variety of memory metrics—such as similarity by cosine (LSA) or context distance (HAL). They are processing models only during concept acquisition (although see Burgess & Lund, 1998; and Li, Burgess, & Lund, in press, for examples of HAL implemented as a processing model).

Thus, GR's experiments illustrate an important limitation of LSA (or HAL) and that is that these models are representational models, not processing models. When one simply uses the vectors like GR did, they are using the representations devoid of processing. This is an appropriate research strategy as long as representational issues are being investigated or if one can directly extrapolate from representation to processing.[2]

However, GR's experiments have as an important concern the nature of the active constructive language-comprehension process. It is simply not reasonable to plop LSA or HAL vectors into a similarity comparison and pretend that it is reflecting the active comprehension process. I think that Landauer and Dumais (1997) are clear about this (see p. 5 of GR). LSA is put forth as a "possible theory about all human knowledge acquisition." A bold claim indeed, but note that the claim is about acquisition, not ongoing language processing. Fur-

[2] An example of this would be the close correspondence between context distance (a representational metric) in HAL and reaction time and semantic priming (reflections of processing) (see Lund & Burgess, 1996; Lund, Burgess, & Atchley, 1995; Lund, Burgess, & Audet, 1996).

thermore, LSA may represent "the underlying mechanism of human cognition." Again note that the claim is that LSA is the underlying mechanism, not the complete set of processes, that are required for active comprehension.

SEMANTIC OR EPISODIC?

Although the LSA and HAL models are completely operationalized, their basic mechanisms break with some cognitive traditions such as the putative relationship between local co-occurrence of words in language and memory organization (Miller & Charles, 1991). Similarity in LSA or HAL is not a matter of matching semantic features; similarity is driven by contextual substitutability. Meaning is represented by the higher order associative contextual history (global co-occurrence)—not just simple co-occurrence information (see Burgess & Lund, 2000; Landauer & Dumais, 1997; Lund, Burgess, & Atchley, 1995; Lund, Burgess, & Audet, 1996). The acquisition process of LSA or HAL captures higher order associations as knowledge structures (not unlike a set of hidden units in a recurrent neural network—see Burgess & Lund, 2000; Burgess, Lund, & Kromsky, 1997).

This is an important point because GR's discussion of associations and LSA and HAL suggest that they may test either episodic instances with their stimuli or ask their participants to engage in complex language-comprehension processing that goes beyond what one might expect from these computational models. LSA and HAL work because of the contextual basis of word meaning acquisition. "Road" and "street" are similar because they occur in similar contexts. Although a person can easily understand a sentence such as *As a substitute for her pillow, she filled up an old sweater with leaves* (from GR Experiment 1), such a sentence poses several problems for contexts models as currently instantiated (a limitation to be sure). Filling a sweater with leaves is a new experience for the model (assuming that the TASA corpus lacks descriptions of this behavior). Thus, this novel episodic experience will not have a vector representation that will be similar to anything within LSA's realm of experience. Beyond this, however, further constructive and integrative processes may be required for the "pillow" meaning to be extracted (or constructed on the fly) from the sweater filled with leaves case. Of course, this is part of GR's point—that LSA and HAL are not capable of this type of processing. Indeed, novel episodic experiences, even though the individual words are in the training corpus, may well not be able to map onto existing representations in the models.

This point is also relevant to the distinction between local co-occurrence (low dimensionality) and global co-occurrence (high dimensionality). For example, the contextual difference between the noun *book* and the verb *booked* (as in *John booked the table*) is substantial. Thus, the local co-occurrence of book with other words will have little, if anything, to do with a representation for the verb booked. In LSA and HAL, local co-occurrence is not important (except at the time of encoding—it is more of an episodic phenomena). What is important for meaning is the global co-occurrence. Global co-occurrence is the weighted collection of local co-occurrences or the context history of a word. Forming hypothesis for experiments based on local co-occurrence or more episodic stimulus constraints and then testing them in a HDM model is almost certain to prove unfruitful.[3]

Meaning in HDM models are a function of their experience and the substitutable nature of words in contexts. To the extent that research questions go beyond these constraints, one is likely to find difficulty in supporting the hypotheses.

[3] Again, the other problem with the innovative verb experiment is that the innovative verbs were never (presumably) encountered by the LSA model during its meaning encoding process in the first place. HDM models have to experience words in meaningful contexts which certainly results in a limitation in the ability to make novel representations. However, as Landauer and Dumais (1997) demonstrate, LSA can learn what a word means long before it is experienced if the model has learned what similar words mean—a striking example of the power of inductive learning.

SPECIFICITY IN A MODEL: TRANSPARENCY VS OPAQUENESS

Whatever LSA and HAL's limitations, clear operationalization is not one of them. Both models are completely transparent and the developers of the models have been very forthcoming in contributing vectors or other data to the research efforts of others. GR's article would not have been possible without the availability of LSA cosines from Landauer's web site (http://lsa.colorado.edu). It is frustrating to read the discussion of GR's experiments and not have a sense of what the operational definitions might be of the important components of indexical theory. One could better consider how LSA or HAL might be tested with respect to indexical theory if there was a transparent presentation of exactly what it means to "mesh in memory." What is an affordance? What are these intrinsic constraints that GR refer to? Exactly how do ideas combine? How does the symbol-grounding process actually take place in indexical theory? The indexical hypothesis does not in any explicit way present the answers to these questions (although cf. Kaschak & Glenberg, 1999). Apparently, affordances provide the wherewithal for memory meshing, then somehow ideas combine, and then presto, almost magically, language comprehension occurs. Without a clearer idea of what these words mean (ironically enough) it is next to impossible to think about what the theory is or what the mechanisms are. Operational definitions need to be spelled out in detail (ideally, to the point to where they could be implemented); one cannot simply afford an operational definition.

This symbol-grounding issue is important in GR's argumentation. In earlier work, we have described how HDM models can make a serious first approximation to dealing with the connection between symbol grounding and representation (Burgess & Lund, 1997b, 2000). The proposal can be made clearly because the models are completely articulated. One problem that embodied theories have is how to explain the representation of abstract words since they are not "symbol-grounded" and are not sufficiently operationalized. The approach to symbol grounding taken by HDM models suggest an obvious solution to the abstract representation issue—the meaning of abstract words can be encoded just like concrete words (see Audet & Burgess, 1999; Burgess & Lund, 1997a). Work on other abstract language issues such as metaphor and idiom is being pursued in my lab (Morrow, Peterson, Burgess, & Eakins, 1999) and also by Kintsch (in press-a) and Landauer (1998).

CONCLUSIONS

Superficially, it is easy to take GR's three experiments at face value. The idea motivating these experiments presented by GR is laudable. They have used LSA to test the idea that certain aspects of language usage can serve as a vehicle to pit a high-dimensional theory of meaning against an embodied theory of meaning. Clearly, GR have detailed some limitations of the high-dimensional approach. GR have demonstrated that if a model does not have appropriate experience, it will not be able to adequately generalize to novel uses of language. They also have clear results that humans can make these novel connections without this experience. This demonstrates that humans are smarter and better language comprehenders than HDM models.

In order for LSA or HAL to reflect the active and creative language-comprehension process, the representations would have to be incorporated into a processing model. Such a model might then simulate the ongoing, more dynamic, activity during comprehension which seems to be the goal of the "on the fly" kinds of novel language usage that was a part of all these experiments. Doing this represents a major research effort to be sure. This "limitation" of the model raises a extremely important question, one which GR discussed in detail. Is this a limitation in principle (at a deep theoretical level), as GR assert, or a limitation in implementation? The art has not reached the state of being able to move beyond speculation on this issue. Neither LSA or HAL are set up to directly deal with the specific language issues that GR raise. Nor is the embodied theory or the indexical hypothesis articulated in anywhere near sufficient form to make predictions without far reaching assumptions as to the linking hypotheses and actual mechanisms. However, GR have framed a set of very important issues that will need to be addressed with future research.

I think it would be wrong to take these experiments at face value for the various reasons discussed earlier. Although I disagree with most of the conclusions that GR derive from their results, I do not think it was unreasonable that these questions were pursued with a set of experiments. Most of the issues discussed in this rejoinder are issues that are frequently raised with HDM modeling efforts. HDM models are “new” in their implementation; however, they have strong roots to the past in cognitive and behavioral psychology (see Burgess & Lund, 2000; and Landauer & Dumais, 1997, for a review). That such controversy can emerge with the models will only to serve to clarify important psycholinguistic issues and encourage the further development of clear theoretical models.

GR have stronger reservations about the scope of LSA (sentences and larger discourse units) than the scope of HAL (words). Summing vector representations to compute higher level meaning may be an unsatisfying approach for some. However, without testing a model that has learned on relevant contextual information it is premature to indict the LSA enterprise, particularly in the context of the wide range of empirical results that have appeared in the literature (Foltz, 1996; Landauer & Dumais, 1997; Wolfe et al., 1998; among many others). No doubt the future will see the development of hybrid computational approaches to sentence and discourse representation. Kintsch (in press-b) has proposed that LSA word vectors could be incorporated into the Construction-Integration theory of comprehension with more ecologically valid results. Another hybrid approach by Miikkulainen and Aguirre-Celis (1999) is incorporating HAL word vectors into a connectionist architecture that develops sentential representations (see Miikkulainen, 1993). Theory development with variants of HDM models will require operationalizing new hybrid approaches such as these. Likewise, theory development with the indexical hypothesis requires this same level of operationalization to seriously compare the models. This is crucial for making clear predictions and for comparing models.

The authors want to make the case that they have shown the soft underbelly of HDM models and conclude that the high-dimensional theory is untenable. Even if one took the experiments by GR at face value, it would be hard to accept the strong conclusions they propose given the broad range of empirical HDM modeling results already in the literature that suggest that these models capture many aspects of word and text meaning. LSA and HAL are context models. Violating the assumption that meaning hinges on contextual experience will guarantee that a context model will not perform well.

REFERENCES

Audet, C., & Burgess, C. (1999). Using a high-dimensional memory model to evaluate the properties of abstract and concrete words. In *Proceedings of the Cognitive Science Society* (pp. 37–42). Hillsdale, NJ: Erlbaum.

Burgess, C., & Conley, P. (1999). Representing proper names and objects in a common semantic space: A computational model. *Brain and Cognition,* **40,** 67–70.

Burgess, C., & Lund, K. (1997). Representing abstract words and emotional connotation in high-dimensional memory space. In *Proceedings of the Cognitive Science Society* (pp. 61–66). Hillsdale, NJ: Erlbaum.

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

Burgess, C., & Lund, K. (1998). Modeling cerebral asymmetries of semantic memory using high-dimensional semantic space. In Beeman, M., & Chiarello, C. (Eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience* (pp. 215–244). Hillsdale, NJ: Erlbaum.

Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 117–156). Hillsdale, NJ: Erlbaum.

Burgess, C., Lund, K., Kromsky, A. (1997). Examining issues in developmental pycholinguistics with a high-dimensional memory model. *Psychonomics Society Abstracts,* **2,** 66.

Foltz, P. W. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments & Computers,* **28,** 197–202.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language,* **43**(3), 379–401.

Kaschak, M. P., & Glenberg, A. M. (1999). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. Manuscript in preparation.

Kintsch, W. (in press-a). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review.*

Kintsch, W. (in press-b). Predication. *Cognitive Science.*

Landauer, T. K. (1998). *How modern computation can turn cognitive psychology into a real science.* Keynote address at the 28th annual conference of the Society for Computers in Psychology (SCiP), Dallas, TX.

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 Bulletin,* **104,** 211–240.

Li, P., Burgess, C., & Lund, K. (in press). The acquisition of word meaning through global lexical co-occurrences. In *Proceedings of the Thirty-First Annual Child Language Research Forum.* Stanford, CA: Center for the Study of Language and Information.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Models, Instrumentation, and Computers,* **28,** 203–208.

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

Lund, K., Burgess, C., & Audet, C. (1996). Dissociating semantic and associative word relationships using high-dimensional semantic space. In *Proceedings of the Cognitive Science Society* (pp. 603–608). Hillsdale, NJ: Erlbaum.

Maxwell, A. (1976). *The Singer enigma.* New York: Popular Library.

Miikkulainen, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory.* Cambridge, MA: The MIT Press.

Miikkulainen, R., & Aguirre-Celis, N. (1999). Personal communication.

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes,* **6,** 1–28.

Morrow, R., Peterson, R., Burgess, C., & Eakins, J. (1999). An initial computational analysis of figurative language: The role of density in idiomatic and metaphoric expressions. *Psychonomics Society Abstracts* **4,** 11.

Wolfe, M., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and texts by Latent Semantic Analysis. *Discourse Processes,* **25,** 309–336.

(Received September 30, 1999)
(Revision received January 31, 2000)