Memory & Cognition
1992, 20 (5), 458-460

Mental models, pictures, and text: Integration of spatial and verbal information

ARTHUR M. GLENBERG
University of Wisconsin, Madison, Wisconsin

and

MARK A. McDANIEL
Purdue University, West Lafayette, Indiana

In the past several years, there has been an acceleration in the publication of cognitive research on the interplay between linguistic and pictorial/spatial information. To report on and encourage this sort of research, we organized a symposium at the 1991 meeting of the Midwestern Psychological Association. The articles in this special section of *Memory & Cognition* are based on the work presented at the symposium. In this introduction, we offer a suggestion for why the integration of linguistic and spatial information is not only a possibility, but a requirement for effective communication. Our suggestion follows the linguistic analysis of the closed-class elements that convey spatial relations, the prepositions (Talmy, 1983). The structure of language provides but a small set of prepositions to encode the vast number of spatial relations that we can perceive. Thus, to understand a situation that a speaker or a writer is conveying, the listener or reader must combine linguistic information with (perhaps metric) spatial information derived from pictures, the environment, or memory.

In 1991, at the meeting of the Midwestern Psychological Association, we held a symposium on the recent surge in research on the relation between language, visuospatial information, and visuospatial modes of communication (pictures, diagrams, maps, etc.). Heretofore, cognitive psychologists have seized on the nominal, surface differences between linguistic and visuospatial presentations as a useful distinction for guiding and circumscribing research on comprehension and memory. This strategy has produced an informative and rich literature on comprehension of connected discourse, and a somewhat less extensive literature on processing and memory for representational drawing, pictures, and maps. Yet the sufficiency of this distinction for creating a fruitful experimental enterprise may not be mimicked in the sufficiency of these modes for individually supporting effective communication.

For instance, consider the assertion by Taylor and Tversky (1992a, p. 495) that "Language is a surrogate for experience." If this were not so, it would be hard to understand how language could be used to inform us about events and objects with which we had no direct contact. Nonetheless, language can be distinctly inferior to experience. Our perceptual apparatus delivers (usually) veridical, organized, detailed representations. Language may not do any of that. In particular, the structure of language is particularly ill-suited for communicating the spatial organization to which our perceptual apparatus seems tuned.

The problem of language and space is revealed by an analysis of closed-class (i.e., limited in size and hard to add to) grammatical terms used to indicate spatial relations. In English, only some 80–100 prepositions are used to convey spatial relations (Jackendoff & Landau, 1991). These are words such as *above*, *across*, *nearby*, and *upon*. Clearly, however, our perceptual apparatus can make much finer metric distinctions. When we wish to use language to convey those finer distinctions, we must often use excessively long locutions that are dependent not on the structure of language, but on culturally agreed upon systems of measure (Jackendoff & Landau, 1991). Thus, the title of this introduction is not just above the abstract, but it is a certain number of millimeters above it, and it is centered on the page a particular number of millimeters from the top, from the sides, and so forth. In English, we simply do not have a term such as *cenbove* (*The title is cenbove the abstract*) to imply that a reference object is not only above, but centered above, a ground (see Jackendoff & Landau, 1991, for their original example).

Talmy (1983) speculates about the origins of this state of affairs. Very specific (but still not complete) specification of spatial relations would require millions of terms. Although this would not in principle be impossible, constraints on natural language militate against the development of such a system. These include the constraints of

Preparation of this article was supported in part by Air Force Office of Scientific Research Grant 89-0367 awarded to A.M.G. Requests for reprints may be sent to A. M. Glenberg, Department of Psychology, 1202 West Johnson Street, Madison, WI 53706, or to M. A. McDaniel, Psychology Department, Purdue University, 1364 Psychology Building, Room 3188, West Lafayette, IN 47907-1364.

learnability, speed of communication, and the capacity of a limited system of phonemes to permit discrimination among the multiplicity of terms required in just this one semantic domain.

Thus, we seem to be faced with a contradiction. Language is a surrogate for experience, but the structure of language prevents it from being an accurate surrogate. The solution is suggested by Talmy (1983) and by much of the research reported at the symposium. Constraints force natural languages to sample the multidimensional space of spatial relations and to ignore possibilities, as in the case of *cenbove*. To fill the interstices, a listener or reader uses spatial information (such as a map or information available from the environment) or prior knowledge regarding spatial organization to construct an image or cognitive model of the situation being described (Glenberg & Langston, 1992; Johnson-Laird, 1983). Thus, the model is the surrogate for the experience, not the language, and the contradiction is eliminated.

This rapprochement suggests that human communication can and will consist of an admixture of language and visuospatial formats (e.g., the mixing of prose and pictures in textbooks, journals, magazines, newspapers, lectures, etc.), presented so that the comprehender can develop a model (understanding) of the experience being communicated. Accordingly, from the vantage point of cognitive psychology, a complete understanding of comprehension and memory processes will depend on our revealing the effects, influences, and representations produced by language and visuospatial formats operating in concert. The research reported in the present issue of *Memory & Cognition* represents efforts to examine various implications and issues emanating from this broad approach. One straightforward idea is that the comprehension and memory of a communication can be augmented if nonlinguistic transmission is exploited, perhaps by assisting in the construction of a model that serves as one product of comprehension. Glenberg and Kruley (1992) focus on one comprehension process that might be enhanced by the inclusion of pictures with text—that of resolving anaphoric reference. In addition, they report robust mnemonic benefits of adding pictures to text. Waddill and McDaniel (1992) show, however, that the mnemonic effects of picture adjuncts can depend on individual differences in reading comprehension. Moreover, in line with the idea that pictures might assist in the construction of a model, pictures designed to target more abstract relational information conveyed in the text (e.g., contrastive relations between two different entities) increased memory for relational information in the text, at least for some readers.

The other papers in this set are concerned with the nature of the representations or models that are created from various input formats (text, maps, drawings). A priori, two kinds of theoretical orientations to this issue can be identified. Because descriptions and drawings are very different on the surface, it might be that they lead to different or separate cognitive representations. An alternative view is that cognitive representations are similar regardless of the media used to communicate. On this view, pictorial (sketches and drawings) and linguistic formats are interpreted into common representations. The authors of the present papers urge the latter view. McNamara, Halpin, and Hardy (1992), using recognition priming, compel the conclusion that people derive an integrated, common representation to store spatial and factual information communicated through maps and language, respectively. Taylor and Tversky (1992) require subjects to learn the information conveyed in a map and to subsequently communicate that information via either a drawing or a description. A key finding is that regardless of whether subjects were required to communicate their knowledge in pictorial or linguistic form, the structure, content, and temporal ordering of the communicated information were similar. Furthermore, these communications paralleled those given by subjects in previous work (Taylor & Tversky, 1992b) who were presented with descriptions rather than maps. This research suggests that comprehenders may be constructing a model of the scenario pictured (or described), from which outputs are similarly generated regardless of the surface format of the output. This is not to say, however, that particular formats did not provide advantages for communicating certain kinds of information. Finally, Denis and Cocude (1992) and Franklin, Tversky, and Coon (1992) investigate in some detail the nature of spatial models constructed from text. Denis and Cocude (1992) suggest that for certain descriptions, much as for maps, the spatial mental model depends heavily on visual imagery. Moreover, their data imply that translating the verbal input into the visual representation is not an all or nothing process but can progress through intermediate stages, stages influenced by the structure of the description itself. Franklin et al. (1992) explore peoples' use of the spatial models (surrogates) derived from text to answer questions about environments conveyed by language. An important point revealed by this research is that although the linguistic description leaves various possible models open, people appear to use preferred constructions.

The health of this area of research may be assessed in several ways. The contributors to the symposium were selected because we knew that their work was relevant to the investigation of interactions between linguistic and spatial formats. Nonetheless, the papers report a wide variety of methodologies, including recall, reading time, priming, drawing, and clustering analysis. Similarly, there is a great diversity in the literature cited. Among the six articles, there are over 20,000 possible pairs of citations. There are only 28 matches, however, and none of the articles cite Jackendoff and Landau (1991) or Talmy (1983), whose ideas we have used to structure this introduction. Although it may be possible to interpret these data as indicating conceptual disarray, we prefer to see them as an index of the exciting variety of problems that are being

tackled and the possibility that the solutions will have a wide impact on cognitive theory.

REFERENCES

DENIS, M., & COCUDE, M. (1992). Structural properties of visual images constructed from poorly or well-structured verbal descriptions. *Memory & Cognition*, **20**, 497-506.

FRANKLIN, N., TVERSKY, B., & COON, V. (1992). Switching points of view in spatial mental models. *Memory & Cognition*, **20**, 507-518.

GLENBERG, A. M., & KRULEY, P. (1992). Pictures and anaphora: Evidence for independent processes. *Memory & Cognition*, **20**, 461-471.

GLENBERG, A. M., & LANGSTON, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory & Language*, **31**, 129-151.

JACKENDOFF, R., & LANDAU, B. (1991). Spatial language and spatial cognition. In D. J. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore festschrift for Lila Gleitman* (pp. 145-169). Hillsdale, NJ: Erlbaum.

JOHNSON-LAIRD, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

MCNAMARA, T. P., HALPIN, J. A., & HARDY, J. K. (1992). The representation and integration in memory of spatial and nonspatial information. *Memory & Cognition*, **20**, 519-532.

TALMY, L. (1983). How language structures space. In H. Pick & L. Acredolo (Eds.), *Spatial orientation: Theory, research, and application* (pp. 225-282). New York: Plenum.

TAYLOR, H. A., & TVERSKY, B. (1992a). Descriptions and depictions of environments. *Memory & Cognition*, **20**, 483-496.

TAYLOR, H. A., & TVERSKY, B. (1992b). Spatial mental models derived from survey and route descriptions. *Journal of Memory & Language*, **31**, 261-292.

WADDILL, P. J., & MCDANIEL, M. A. (1992). Pictorial enhancement of text memory: Limitations imposed by picture type and comprehension skill. *Memory & Cognition*, **20**, 472-482.