

Embodied Cognition, Perceptual Symbols, and Situation Models

Rolf A. Zwaan
Department of Psychology
Florida State University

It is difficult to see how current models of discourse comprehension can be "scaled up" to account for the rich situation models that may be constructed during naturalistic language comprehension, as when readers are immersed in the story world. Recent proposals about embodied cognition and perceptual symbols, such as those put forth by Glenberg and Robertson and by Roth might offer a framework to address this problem more successfully than traditional approaches that rely on amodal symbol systems.

Suppose I tried to explain to you how to ride your mountain bike over a log, 16 in. in diameter, that has fallen across the trail. Here is what I might say:

Make sure you are in a gear that is not too high and not too low. Make sure you have sufficient speed and that you are perpendicular to the log. Then get up out of the saddle, keep your feet still, and lift the front wheel so that it clears the log. Your large gear will hit the log and your momentum will carry you across the log. Stretch your arms before the front wheel hits the ground so that you don't flip over. Start pedaling as soon as your front wheel hits the ground.

Readers without mountain-biking experience will be able to construct a situation model that probably includes a protagonist on a bike, the starting position and orientation of the bike relative to the log, the position and orientation of the log relative to the trail, the temporal sequence of actions, and the protagonist's goal. Evidently, such a situation model is not sufficient to have an individual safely negotiate the log. (So do not try this at home.) However, more to the point, compared to novices, people with mountain-biking experience will be able to construct a far richer situation model of this sequence of events.

What are the additional components an experienced mountain biker might be able to add to achieve a higher resolution situation model? Here are some candidates: proprioceptive, motor, optic flow, and haptic information. For example, the expert might be able to mentally simulate the positions of his or her limbs—how the knees are arched and how the feet are positioned relative to each other. The expert will also have some sense of the amount of strength needed to lift the front wheel. The expert might have some sensation of the speed at which the trees along the trail pass by (optical flow) and the speed at which the log approaches. Finally, the expert might “feel” how his or her hands are curved around the handles and the pressure of the pedals on his or her soles.

Note that this information is not just ornamental. For example, the optic flow information will tell the expert when to start pulling the handle bar. Proprioceptive information will tell the expert when the front wheel has been raised high enough (maybe because of the relative position of the hands to the center of gravity). Thus, this information helps construct a temporally and spatially much finer grained situation model. Moreover, experts and novices might differ with respect to their perspective on the situation. Rather than a first-person perspective, novices might use a “third-person” perspective to represent the situation: viewing someone else, from the side, on a mountain bike.

Recent theoretical proposals involving embodied cognition and perceptual symbols (e.g., Barsalou, 1999; Glenberg, 1997; MacWhinney, 1999) provide a conceptual framework in which such matters can be better investigated than in more traditional “amodal” symbolic frameworks. In amodal symbol systems, there is no analog relation between the referent and the mental representation. Instead, the relation is arbitrary. Perceptual symbols, on the other hand, have an analog relation with their referent (Barsalou, 1999) and therefore provide more information about the referent than do amodal symbols (Glenberg & Robertson). As such, modal symbol systems might provide a fruitful framework for studying situation-model construction in text comprehension. The two target articles in this issue (Glenberg & Robertson; Roth) provide some intriguing initial empirical evidence for this new view. The purpose of this commentary is to briefly examine some of the implications of an embodied–perceptual-symbol view for theories of situation models.

WHAT IS REPRESENTED IN SITUATION MODELS?

Situation models are mental representations of the situations described in language (Kintsch, 1998; Zwaan & Radvansky, 1998). Intuitively, when we read a novel, we feel that we construct a rich mental representation of the story world. We feel “immersed” in the story world. We have some idea of what the protagonist’s environment looks like and—if we are familiar with the environment—we may

even be able to move along with the protagonist through it. We keep track of the protagonist's goals and his or her successful and failed attempts to realize them. We often also infer physical causation. For example, we mobilize our knowledge of water and fire to infer that the fire went out because someone poured water on it. In addition, we recruit our rich knowledge of human emotions to infer that the protagonist is frustrated when his or her goal is not achieved. We are caught up in the temporal sequence of events such that those events that are close to us in the story world are more accessible in our memory than events that happened some time ago. Situation-model researchers have amassed solid evidence for each of these claims.

However, the phenomenological experience of being immersed in a story world goes far beyond this. When reading a story, we may "experience" cold wind blowing in our face, the smell of stale beer, a kiss on our lips, and a hot piece of pizza sticking to the roof of our mouth. This may seem far-fetched, but it is not. Listening to a scenario about people smoking after-dinner cigarettes has been shown to elicit physiological and cognitive responses indicative of smoking urges in smokers (e.g., Drobles & Tiffany, 1997; Zwaan & Truitt, 1998). In fact, scenarios like these are as efficient in eliciting smoking urges as are *in vivo* cues, such as holding (but not smoking) a lit cigarette. It is not unreasonable to hold situation-model theories accountable for explaining this kind of vicarious experiencing by way of language. However, current theories and models fall far short of meeting this expectation.

Discourse psychologists have obtained a rather good understanding of several aspects of situation-model construction. For example, we know a great deal about how readers keep track of protagonists' goals and plans (Graesser, Singer, & Trabasso, 1994). We are also beginning to understand how readers keep track of time and the conditions of how and when they keep track of space. However, it is a little more difficult to see how amodal symbolic models could account for the full complexity of discourse comprehension. It would seem that mentally representing essential characteristics of situations, such as perspective, motion, and emotions, would involve an overwhelming number of amodal propositions. The question is not so much, then, whether symbolic systems can represent these structures and processes because by definition they can. The question is much more whether they can do so given what we know about the limitations of working memory and attention plus what we know about language processing. For example, at a normal reading-for-pleasure rate of 400 words per minute (Nell, 1988), is it realistic to assume that people will construct and integrate thousands of propositions to construct a high-resolution situation model?

There have been recent attempts to integrate perceptual processes with higher level cognitive processes in symbolic systems. For example, a new version of Anderson's Adaptive Control of Thought-Revised (ACT-R) theory is called ACT-R/PM, in which PM stands for perceptual motor (Anderson & Lebiere, 1998). This model does a very nice job simulating participants' performance on

the Sperling task and on visual search tasks. However, it is difficult to see whether this model will scale up to more complex and naturalistic tasks. As Anderson and Lebiere noted: "The one difference between an ACT-R simulation and a human is that ACT-R's whole world is the computer screen, the mouse, and the keyboard, whereas this is only a small part of the subject's world" (p. 144). How big this difference is, is of course the main bone of contention. Also, one might wonder how similar the participants' and computer's representations of the monitor, mouse, and keyboard really are. For example, unlike a human, the computer's performance would not be affected by a glare on the screen or the awkward shape of the computer mouse (such as the one that comes with the iMac).

Following Barsalou (1999), Glenberg and Robertson argue that perceptual symbols might provide a solution to the problem of accounting for the full complexity of discourse comprehension because they would make perceptual information available very quickly and presumably at little processing cost.

The challenge posed by the two target articles in this issue has two prongs. First, there is the argument that situation models involve more than declarative knowledge. Second, there is the argument that situation models are not amodal representations, that is, representations that are not based on a conventional symbol system. Such a system is conventional in the sense that it is grounded in arbitrary rather than analog relations between symbols and their referents. I discuss these arguments in turn and review their theoretical merits as well as the empirical support they have received in the research in the two target articles.

SITUATION MODELS INCLUDE MORE THAN DECLARATIVE KNOWLEDGE

Glenberg and Robertson as well as Roth argue that people use more than just declarative knowledge when representing situations. Earlier studies are consistent with this view. Klatzky, Pellegrino, McCloskey, and Doherty (1989) found that the comprehension of verbally described actions is facilitated when participants are allowed to first form their hand into a shape appropriate for the action (e.g., fingers pinched together when reading about throwing a dart).

Morrow and Clark (1989) obtained interesting evidence suggesting that readers' interpretation of motion words, such as *approach*, depends on physical characteristics of the situation. Specifically, it depends on the size of the object and the landmark. For example, in "The mouse approached the fence" participants represented the mouse as being much closer to the fence than they did the tractor in "The tractor approached the fence." Similarly, the tractor would be closer to the landmark in "The tractor approached the fence" than in "The tractor approached the farm." It seems that an embodied-perceptual account could explain this more readily than an amodal symbolic system. For how would the various meaning gradations of action verbs be represented in such a system? It would

seem unlikely that we have stored multiple copies of *approach* in which its meaning is defined relative to specific combinations of objects and landmarks.

Franklin and Tversky (1990) showed that, when comprehenders have constructed a situation model based on a story that places them in the middle of a location, they behave as though they were physically in that situation. For example, objects in front of them are more accessible than objects behind them.

Thus, there is an emerging pattern of findings suggesting that the way we interact with and observe our environment informs our understanding of verbal materials. To put it differently, there is a pattern of evidence suggesting that the way we build situation models when reading text is derived from how we build situation models of our environment. To be sure, a great deal of procedural knowledge can be successfully modeled by amodal symbol systems (Williams, Hultman, & Graesser, in press). However, the target articles invite us to wonder whether the best way to understand situation models is in terms of abstract pseudolinguistic concepts such as amodal propositions.

AMODAL VERSUS MODAL PROPOSITIONS

As said, a vast number of amodal propositions would have to be involved to account for the richness of situation models, such as the mountain-biking expert's situation model of how to get over the log. Thus, it might not be impossible to construct an amodal situation model, it might just be psychologically implausible. To decide this issue, empirical evidence is needed. The question is whether the target articles have provided such evidence. It is easiest to examine this question with respect to Glenberg and Robertson's study because theirs is an experimental one. It would seem that one could come up with a small set of amodal propositions that distinguishes Glenberg and Robertson's listen and index condition from the other condition. Each time the hand in the video points to a particular component of the compass, the viewer could form a (complex) proposition such as "The arrow is the red triangle at the top of the compass." Thus, an amodal symbolic system might explain this evidence (post hoc!) rather easily. However, as Barsalou (1999) pointed out, explaining a finding post hoc is really not as impressive as both predicting and explaining it. Furthermore, the mountain-biking example shows that large and maybe unmanageable numbers of amodal propositions might be needed to represent complex situations, specifically situations involving motion.

It is important to ask whether perceptual-symbol systems can account for the same findings amodal symbolic systems can. After all, one does not want to throw out the baby with the bath water. Barsalou (1999) argued that perceptual symbols can exhibit all the features of propositions. Kintsch (1998) made a strong case for the role of amodal propositions in discourse comprehension. He cited evidence from three sources pointing to the existence of propositions: cued recall, reading times, and priming. For example, Kintsch discussed cued-recall patterns

for sentences such as "The mausoleum that enshrined the czar overlooked the square." When participants are presented with the cue *overlook*, they are much more likely to recall *square* than *czar*, even though *czar* and *overlook* are neighbors in the sentence's surface structure. Kintsch analyzed this sentence into two propositions: OVERLOOK[MAUSOLEUM, SQUARE] and ENSHRINE[MAUSOLEUM, CZAR].

It would seem that a perceptually based, embodied explanation could arrive at a similar parsing but in a decidedly different way. Using the framework proposed by MacWhinney (1999), the sentence could be parsed as follows. The mausoleum overlooking the square is an object-centered frame, in which the reader takes the perspective of the object, the mausoleum, that overlooks the square. The mausoleum now becomes what MacWhinney calls a secondary human perspective. The mausoleum enshrining the czar would entail a perspective shift in which the reader takes the perspective of the czar's corpse, and the mausoleum becomes an external object. Consequently, when participants are presented with the cue *overlook*, they should be more likely to recall *mausoleum* than *czar*.

As this example shows, analyses based on perceptual-embodied frameworks might yield the same set of constituents as analyses based on an amodal propositional framework. Obviously, much more extensive analyses are needed before such a determination can be made with confidence. However, the example serves to make two important points. First, perceptual-embodied analyses are fundamentally different from amodal symbolic analyses. Second, despite this, the two types of analyses may yield identical predictions in many cases. This is problematic from an empirical point of view. Thus, it is critical to try to generate contrasting and empirically testable hypotheses.

A complication is that the two frameworks may not have to be mutually exclusive. Perhaps both are used. It may depend on the depth of our background knowledge to what extent our representations are embodied. As mentioned earlier, mountain-biking novices are still able to form situation models based on the description on how to negotiate a log across the trail—just as participants were still able to form some situation model when reading Bransford and Johnson's (1972) famous "washing clothes" passage, even without having seen the title. Thus, the level of resolution of a situation model could range from very coarse—involving something like semantic primitives (e.g., "ingest food")—to very detailed.

Proponents of an embodied view argue that, even in such instances, embodied representations play a role (e.g., Barsalou, 1999; see also Roth). Although this is possible, it does not rule out a priori that amodal representations are also involved. Detailed representations might involve embodied information as well as amodal representations. One can envisage many circumstances in which comprehenders acquire some understanding even though they cannot index any of the terms used in the text. The washing clothes episode would be a good example of this. In this case, comprehenders probably do not have a sufficiently detailed representation to carry out the action. However, they do have some situational

understanding. For example, they do know the chronological order of the events, even though they do not know precisely what those events are. A more detailed situation model can be built if we know what the referents of these events are. For example, this might provide us with more temporal and spatial resolution of the sequence of events. This representation could, in turn, be enhanced when we could appeal to direct experiences with washing clothes.

CONCLUSIONS

The target articles provide empirical evidence for the role of perceptual information in situation-model construction. However, on one hand, it could be argued that neither article provides direct empirical evidence that this information is represented analogically. On the other hand, some of the theoretical considerations put forth by the authors and by others (e.g., Barsalou, 1999; MacWhinney, 1999), as well as evidence from related studies, such as those by Franklin and Tversky (1990), Klatzky et al. (1989), Morrow and Clark (1989), and Glenberg and Robertson (in press), suggest that the time is ripe for investigations into the role of perceptual representations in discourse comprehension. Such investigations could provide a major impetus for situation-model theories, which, in their current form, may have a difficult time accounting for the phenomenological richness of discourse comprehension.

ACKNOWLEDGMENTS

I thank Art Glenberg and Art Graesser for comments on a previous version of this article.

REFERENCES

- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717-726.
- Drobes, D. J., & Tiffany, S. T. (1997). Induction of smoking urge through imaginal and *in vivo* procedures: Physiological and self-report manifestations. *Journal of Abnormal Psychology*, 106, 15-25.
- Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, 119, 63-76.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1-19.

- Glenberg, A. M., & Robertson, D. A. (in press). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371–395.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.
- Klatzky, R. L., Pellegrino, J. W., McCloskey, B. P., & Doherty, S. (1989). Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. *Journal of Memory and Language*, 28, 56–77.
- MacWhinney, B. (1999). The emergence of language from embodiment. In B. MacWhinney (Ed.), *The emergence of language* (pp. 213–256). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Morrow, D. G., & Clark, H. H. (1989). Interpreting words in spatial descriptions. *Language and Cognitive Processes*, 3, 275–291.
- Nell, V. (1988). The psychology of reading for pleasure: Needs and gratifications. *Reading Research Quarterly*, 23, 6–50.
- Williams, K. E., Hultman, E., & Graesser, A. C. (in press). CAT: A tool for eliciting knowledge on how to perform procedures. *Behavior Research Methods, Instruments, & Computers*.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–185.
- Zwaan, R. A., & Truitt, T. P. (1998). Smoking urges affect language processing. *Experimental and Clinical Psychopharmacology*, 6, 325–330.