Research Article

THE EFFECT OF IMPLIED ORIENTATION DERIVED FROM VERBAL CONTEXT ON PICTURE RECOGNITION

Robert A. Stanfield and Rolf A. Zwaan
Florida State University

Abstract—*Perceptual symbol systems assume an analogue relationship between a symbol and its referent, whereas amodal symbol systems assume an arbitrary relationship between a symbol and its referent. According to perceptual symbol theories, the complete representation of an object, called a simulation, should reflect physical characteristics of the object. Amodal theories, in contrast, do not make this prediction. We tested the hypothesis, derived from perceptual symbol theories, that people mentally represent the orientation of an object implied by a verbal description. Orientation (vertical-horizontal) was manipulated by having participants read a sentence that implicitly suggested a particular orientation for an object. Then recognition latencies to pictures of the object in each of the two orientations were measured. Pictures matching the orientation of the object implied by the sentence were responded to faster than pictures that did not match the orientation. This finding is interpreted as offering support for theories positing perceptual symbol systems.*

A classic debate in cognitive psychology and cognitive science has concerned the format in which information is stored and manipulated in the human brain. The historically prevalent theory of knowledge representation in cognitive science has been that of the amodal or (propositional) symbol system (e.g., Fodor, 1975; Kintsch, 1998; Newell & Simon, 1972; Pylyshyn, 1981, 1984). Recently, however, theorists have presented a potentially viable alternative to amodal systems in the form of perceptual symbol systems (Barsalou, 1999a, 1999b). There has not yet been a great deal of empirical research that directly tests the contrasting predictions generated by amodal and perceptual symbol systems. The purpose of this article is to make such an attempt.

It is informative to acknowledge a problem that researchers face when trying to determine which of these competing theories is probably more correct. The problem is that amodal systems are virtually unfalsifiable (Anderson, 1978; Barsalou, 1999a, 1999b). Amodal symbol systems can explain anything, albeit post hoc, that perceptual symbol systems can. This can be accomplished by construing an ever-increasing network of amodal propositions that possess an ever-increasing number of operations capable of being performed on those propositions—amodal systems provide no theoretical limits on the numbers of propositions or operations. Because of this massive explanatory power that amodal systems possess, any attempt to falsify them might be doomed to fail. Thus, it seems that the most convincing falsification would arise not from finding effects that amodal systems cannot explain, but instead from finding effects that (a) amodal systems do not predict and (b) perceptual symbol systems explain more elegantly (i.e., more parsimoniously) (Barsalou, 1999b). As Barsalou argued, the theory that has the ability to predict rather than postdict effects and is more parsimonious should be favored.

Address correspondence to Robert A. Stanfield or Rolf A. Zwaan, Department of Psychology, Florida State University, Tallahassee, FL 32306-1270; e-mail: stanfiel@psy.fsu.edu or zwaan@psy.fsu.edu.

One possible method of empirically differentiating between these two systems is by closely examining their most important theoretical difference, namely, the proposed relationship between the internal symbol and the external referent. This relationship is characterized as being arbitrary and linguistic-like in amodal systems, whereas it is characterized as being analogue and perceptual in perceptual systems (Barsalou, 1999b). As an example, consider the potential mental representation for a dining room table. In an amodal system, such a representation would be something as arbitrary as the word *table* itself. Note that this representation is arbitrary because there is no a priori reason why the letter string t-a-b-l-e should be used to represent the object Americans call a table. In principle, it could just as easily be called *mesa*. The label is merely an arbitrary product of an English-speaking society. However, the perceptual symbol system's representation of the table is not arbitrary. Instead, the representation (if it could be viewed) would look something like the table, with a flat surface and four legs (or a pedestal, depending on the type of table), and, barring cultural differences, a person's representation of the table would be similar regardless of whether he or she is Korean, Egyptian, English, or American.

The analogue relationship between symbol and referent in perceptual symbol systems means that changes in the referent will cause changes in the perceptual symbol (Barsalou, 1999b). For example, if the referent changes orientation (e.g., if the table is upside down on the floor), so too will the representation, whereas this would not occur in an amodal system. It is important to note that the perceptual representation for a table would not be created from a single perceptual symbol, but would instead be defined by the combination of a series of perceptual symbols for different components of the referent (e.g., top, legs, color, shape, wood grain), and orientation is only one of these symbols. The single unified representation of an object that is produced by the combination of perceptual symbols is called a *simulation* (see Barsalou, 1999b). A simulation occurs when the constituent perceptual symbols for a given object are meshed together (Glenberg, 1997). This meshing process is guided, in turn, by the affordances (see Barsalou, 1999b; Gibson, 1979; Glenberg, 1997; Glenberg & Robertson, 1999, 2000; MacWhinney, 1998) of the symbol or object. Affordances are defined here as simply the properties associated with an object relative to an observer or user. Dining room tables, for example, afford being turned upside down by humans but not by cats, whereas mountains do not afford being turned upside down by either humans or cats.

Thus, theories involving perceptual symbol systems predict that changes in the physical referent (e.g., table) would cause similar changes in the perceptual symbols that produce the simulation of the referent. This in turn leads to predictions about the interaction between mental representations and their referents. Specifically, perceptual symbol theories predict that the output of simulations should mimic the output of operations performed during actual perception. In other words, a simulation should share certain "physical" characteristics with its real-world counterpart. If the referent grows taller, so too should the simulation. If the referent changes color, so too should the simulation. Any transformation the referent undergoes should cause an analogous transformation in the simulation. Moreover, and most central to the current study, as Barsalou (1999b) argued, these effects should extend to simulations formed by the creation and integration of situation models (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998).

Barsalou (1999b) provided a useful illustration of this point. He predicted that if an individual reads a sentence such as "John put the pencil in the cup," then, because of the affordances associated with pencils and cups, the simulation of that sentence should include a vertical orientation for the pencil, just as if one were to actually experience it in such a position. This is so because one affordance of a pencil is that it is usually long, and to simulate placing it in a cup, one must place it in vertically, because the "long" affordance prevents doing it horizontally. However, the simulation of a sentence like "John put the pencil in the drawer" should include a horizontal orientation for the pencil, for a similar reason. Note that although information about orientation is a natural by-product in the simulation of these examples in a perceptual system, this is not the case in an amodal system. Given the sentence "John put the pencil in the cup," a typical amodal representation would be constructed as

PUT(JOHN, PENCIL, CUP)

Though one could postulate an additional argument describing the pencil's orientation, such an inference would not be predicted by the assumptions common to amodal systems.

It is important to note that this argument does not assume that all possible affordances of various perceptual symbols are activated during a simulation; rather, only those relevant to comprehension are activated. For example, in examining the role of affordances in the understanding of novel denominal verbs (i.e., verbs formed from nouns, as in "He hammered the nail"), Kaschak and Glenberg (2000) found that the affordance most relevant to the comprehension of a given passage was verified faster than an affordance not relevant to comprehension. Most important, Kaschak and Glenberg varied the passages so that one affordance was more relevant than another in one condition (and thus was verified faster), but the relevance of the two affordances (and consequently the speed with which they were verified) was reversed in another condition. In the current example, orientation is relevant because the "length" affordance limits the ways in which the sentence can be successfully simulated.

Unfortunately, it is not unproblematic to directly assess the output of a mental simulation to test the predictions generated by theories positing perceptual symbol systems. However, the analogue nature of perceptual symbol systems allows us to expand on Barsalou's idea one step further and argue that, just as changes in the referent produce similar alterations in the simulation, so too would changes in the simulation have implications for the interpretation of the referent. Specifically, one could predict that verification times for a picture of a pencil in a vertical orientation should be shorter than verification times for a picture of a pencil in a horizontal orientation given the sentence "John put the pencil in the cup," and that the relative verification times should reverse given the sentence "John put the pencil in the drawer." This is to be expected because of the perceptual nature of the simulations. Just as actually seeing an object would facilitate recognition of that object later, so too should simulating the object facilitate later recognition of the object. The greater the overlap between the object and the simulation (i.e., the more features they have in common), the greater the degree of facilitation.

In the experiment we report here, we examined the effect of a referent's orientation on the recognition latencies to a picture of that referent when orientation was made relevant by the previous sentence. Participants viewed a series of sentences, each followed by a picture, and answered "yes" or "no" to indicate whether the pictured object had been mentioned in the preceding sentence. The orientation of the objects was varied so that it either matched or did not match the orientation implied by that sentence. Under these conditions, perceptual symbol theories predict that participants will exhibit shorter recognition latencies to matching pictures than to mismatching pictures. Amodal theories, in contrast, predict that recognition latencies will not differ between these conditions.

The current experiment also examined the potentially mediating effect of individual differences in mental representations, by incorporating the Flags test (Thurstone & Jeffreys, 1956) of general spatial ability as a covariate. We used the Flags test because it is a mental rotation task, and mental rotation is close to what participants might do in recognizing the pictures.

METHOD

Participants

Forty undergraduate students enrolled in introductory psychology courses at The Florida State University participated to receive credit toward a course requirement.

Materials

One hundred and five black-and-white line drawings were used in this experiment. Most were obtained from the Snodgrass and Vanderwart (1980) normed pictures, but some pictures from a popular clip-art program were used as well. Eighty-one pictures served as filler items, and 24 (non-semantically related) pictures served as experimental items. Each of the 24 experimental items was rotated 90° on its vertical axis, to produce a total of 48 possible experimental items, half with a vertical orientation and half with a horizontal orientation. Each picture was scaled to occupy a square measuring 3.5 × 3.5 in.

It was important to ensure that picture recognition would not be affected by the canonicality or view specificity of the pictures (Christou & Bülthoff, 1999; Dror & Kosslyn, 1998; Kosslyn et al., 1994; Lawson & Humphreys, 1996). To that end, we pilot-tested the materials with 24 participants. They were presented with an object name and then a picture, and had to decide if the picture matched the name. Each participant saw each of the items once, with orientation of the experimental items counterbalanced across groups. Half of the items required a "no" response and half required a "yes" response. All experimental items required a "yes" response. After they made the yes/no decision for each picture, participants were asked to rate the general quality of the picture on a 7-point Likert scale (1 = *poor quality*, 7 = *high quality*). In order to be used in the current experiment, an item had to meet the following conditions: First, a two-tailed paired-sample t test did not reveal a significant or marginally significant difference between the response times for the horizontal and vertical orientations (i.e., $p > .10$). Second, the average mean difference between response times to the horizontal and vertical orientations was not greater than 2 standard deviations above the average mean difference for all items. Third, the picture received a rating above 4.0 on the 7-point scale. Fourth, the median response time for the picture was less than 1,200 ms. The only criterion for filler items was that they were all rated above 4.0 on the Likert scale.

We created 129 sentences to accompany the pictures: 81 filler sentences, 24 vertical-orientation sentences, and 24 horizontal-orientation sentences. Fifty-three of the filler sentences did not mention the pictured object and would thus require "no" responses. We attempted to control for the position of the object noun in the sentence by placing the object noun at the beginning, in the middle, and at the end of the sentence in approximately equal numbers of sentences. All sentences, even those requiring a "no" response, contained at least one concrete noun.

It was important to ensure that the two sets of experimental sentences (horizontal, vertical) equally constrained the potential orientations of their respective objects. In other words, it was undesirable to have a vertical object be a plausible representation of a horizontal sentence, and vice versa. We therefore conducted a four-alternative forced-choice experiment in which another 24 participants were presented, on each trial, with one of the experimental sentences and the corresponding horizontal and vertical pictures. The participants' task was to choose among the following alternatives: (a) the vertical picture best matched the sentence, (b) the horizontal picture best matched the sentence, (c) the two pictures equally matched the sentence, and (d) neither picture matched the sentence. Participants matched horizontal pictures to horizontal sentences on average 5 out of 8 times and matched vertical pictures to vertical sentences on average 5.5 out of 8 times. This difference was not significant, $t(23) = 0.35$, $p > .1$.

The sentence-picture verification task was presented using a PowerMac 7100/66 AV with a 13-in.-viewable Apple AudioVision 14 display using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). The Flags test was administered with paper and pencil.

Design and Procedure

Four lists of sentence-picture pairs were created. A given sentence-picture pair could occur in one of four conditions: vertical-vertical, horizontal-horizontal, vertical-horizontal, and horizontal-vertical. Each participant was exposed to each condition, but each item appeared in only one condition per list, and each participant saw only one list. This design produced a 2 (match vs. mismatch) × 2 (picture orientation) × 4 (list) design, with match and orientation as within-participants variables and list a between-participants variable.

The experiment started with five practice trials to familiarize the participants with the task. During each trial, participants first saw a left-justified and vertically centered fixation point for 250 ms; then, a sentence appeared. Participants pressed the space bar when they had understood the sentence, another fixation point appeared in the center of the screen for 250 ms,[1] and then a picture appeared. Participants then had to determine if the pictured item had been mentioned in the previous sentence, using the period (".") key for "yes" responses and the "x" key for "no" responses. Participants were instructed that reaction times were being measured and that it was important for them to make the decisions about the pictures as quickly as possible. They were also told that, at various points throughout the experiment, they would be asked to recall the sentence they had just read, so they should be sure to read each sentence carefully. Half of all filler items included a sentence recall task after the verification decision. The first half of the experiment took approximately 35 to 40 min to complete. After finishing the computerized task, participants were given the Flags test. The entire procedure took approximately 45 to 50 min.

RESULTS AND DISCUSSION

Incorrect responses were dropped from the analyses; less than 5% of the data was excluded for this reason. The median correct response time per condition per participant was used in the analyses in order to decrease the effects of extreme outliers.

Results were analyzed using a 2 (match vs. mismatch) × 2 (horizontal vs. vertical) × 4 (list) within-participants analysis of covariance (ANCOVA), with score on the Flags test the covariate of interest. List was treated as a factor in order to increase the power of the analysis by eliminating error due to random pairings of item to condition (Pollatsek & Well, 1995; Raaijmakers, Schrijnemakers, & Gremmen, 1999), but given its lack of theoretical interest, it is not be discussed further.

There was no main effect of orientation, $F < 1$, nor a reliable interaction between orientation and match, $F(1, 36) = 1.94$. However, participants responded more quickly ($M = 838$, $SD = 331$ vs. $M = 882$, $SD = 329$) when the picture matched the implied orientation of the sentence than when it did not, $F(1, 36) = 6.36$, $p < .05$.

There was also a significant effect for the Flags test, $t(36) = -2.54$, $p < .05$. A median-split analysis was performed on Flags data to determine the nature of the effect. The Flags test did not interact with match, $F < 1$, or orientation, $F(1, 32) = 1.45$, nor was there a significant Flags Test × Match × Orientation effect, $F < 1$. However, higher spatial ability ($M = 72$; range: 21–118) was associated with shorter response latencies in both the match ($r = -.33$, $p < .05$) and the mismatch ($r = -.38$, $p = .01$) conditions, suggesting that individuals with better spatial ability responded faster than individuals with poorer spatial ability. The Flags data suggest that perceptual symbols are used by all individuals regardless of their spatial competency, and are not limited to individuals with good spatial cognition, although it should be noted that the analyses may not have been powerful enough to detect differences between high- and low-ability participants.

GENERAL DISCUSSION

As predicted, the recognition of an object mentioned in a sentence was influenced by the orientation of the object. Specifically, recognition latencies were shorter for pictures of objects whose orientation matched the orientation suggested by the prior text than for pictures of objects whose orientation did not match the one suggested by the prior text. This finding is in line with the predictions of perceptual symbol systems (Barsalou, 1999b), but not with those of amodal symbol systems.

However, it could be argued that because participants saw a picture after each sentence, they might have been encouraged to adopt a strategy in which they purposefully generated an image to be compared with the picture, and that the obtained results are an artifact of this strategy (e.g., Glucksberg, Trabasso, & Wald, 1973). We believe this is not the case for several reasons. First, a task analysis suggests that our procedure would discourage participants from adopting such a strategy. The sentences typically (61%) described more than one object. Thus, participants would have had to image all objects while knowing

1. An interstimulus interval of 250 ms was used because of research by Lawson and Humphreys (1996) suggesting that exposure to long interstimulus intervals may give participants time to access more view-invariant representations of the objects, which could possibly mask the effects of interest in the current experiment. Pilot testing seemed to confirm this assumption.

only one would be shown. Furthermore, half of the pictures were not mentioned in the preceding sentence. Thus, participants would have been doing a great deal of imaging with very little payoff. Second, in a related study (Stanfield, 2000) using the same materials and procedure, only 25% of the participants reported trying to actively generate images. Third, if participants had used an image-generation strategy, then the Flags test should have had a greater influence on the results, as individuals with better spatial cognition would be more likely and more able to generate such images than individuals with poorer spatial cognition. However, as mentioned, the pattern of results was the same for participants with low and high spatial ability.

A fourth reason is that if participants had used this strategy, the obtained effect should have been greater in later trials as opposed to earlier ones, given that it would take exposure to multiple trials to discover and implement the strategy. However, a follow-up analysis showed that there was a 51-ms effect in the first half of the experimental trials and a 45-ms effect in the second half, suggesting that the effect was present from the onset of the experiment.

Finally, an image-generation strategy is not sufficient to explain the obtained results. One must wonder why individuals would form integrated images of the sentences when on half the experimental trials the object would not be in the proper orientation, and how individuals would form integrated images of the sentences if not as the result of a simulation.

The idea that comprehenders have access to perceptual information during reading has important consequences for research on situation models (Zwaan & Radvansky, 1998). It has already been demonstrated that comprehenders have access to spatial and temporal information relevant to the protagonist (e.g., Morrow, Greenspan, & Bower, 1987; Zwaan, Madden, & Whitten, 2000), but the current research suggests similar information might be available about all objects contained in the situation model.

Although the current experiment cannot be taken as an unconditional falsification of amodal symbol systems (but for other evidence supporting perceptual symbol systems, see, e.g., Glenberg & Robertson, 1999; Glenberg, Robertson, Jansen, & Johnson-Glenberg, 1999), it does suggest that amodal theories, as they stand now, are insufficient to fully explain comprehension and that, moreover, perceptual symbol systems or, at the very least, hybrid systems (e.g., Harnad, 1990; Paivio, 1971) are a viable alternative.

Acknowledgments—This research is based in part on the first author's master's thesis under direction of the second author. We thank Neil Charness and John Kline for helpful comments on earlier versions of this article. We also thank Peisha Walker for assistance with data collection.

REFERENCES

Anderson, J.R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85*, 249–277.

Barsalou, L.W. (1999a). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, *28*, 61–80.

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

Christou, C.G., & Bülthoff, H.H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition*, *27*, 996–1007.

Cohen, J.D., MacWhinney, B., Flatt, M., & Provost. J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*, 257–271.

Dror, I.E., & Kosslyn, S.M. (1998). Age degradation in top-down processing: Identifying objects from canonical and noncanonical viewpoints. *Experimental Aging Research*, *24*, 203–216.

Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University 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.

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*, 379–401.

Glenberg, A.M., Robertson, D.A., Jansen, J.L., & Johnson-Glenberg, M.C. (1999). Not propositions. *Cognitive Systems Research, 1*, 19–33.

Glucksberg, S., Trabasso, T., & Wald, J. (1973). Linguistic structure and mental operations. *Cognitive Psychology, 5*, 338–370.

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

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

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

Kosslyn, S.M., Alpert, N.M., Thompson, W.L., Chabris, C.E., Rauch, S.L., & Anderson, A.K. (1994). Identifying objects seen from different viewpoints: A PET investigation. *Brain*, *117*, 1055–1071.

Lawson, R., & Humphreys, G.W. (1996). View specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 395–416.

MacWhinney, B. (1998). The emergence of language from embodiment. In B. MacWhinney (Ed.), *The emergence of language* (pp. 213–256). Mahwah, NJ: Erlbaum.

Morrow, D.G., Greenspan, S.L., & Bower, G.H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language*, *26*, 165–187.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.

Pollatsek, A., & Well, A.D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 785–794.

Pylyshyn, Z.W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88*, 16–45.

Pylyshyn, Z.W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: Bradford Books.

Raaijmakers, J.G., Schrijnemakers, J.M.C., & Gremmen, F. (1999). How to deal with "The language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, *41*, 416–426.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.

Stanfield, R.A. (2000). *The effects of verbal context on picture recognition.* Unpublished master's thesis, Florida State University, Tallahassee.

Thurstone, L.L., & Jeffreys, T.E. (1956). *Flags: A test of space thinking*. Chicago: Chicago Industry Service.

van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Zwaan, R.A., Madden, C.J., & Whitten, S.N. (2000). The presence of an event in the narrated situation effects its activation. *Memory & Cognition, 28*, 1022–1028.

Zwaan, R.A., & Radvansky, G.A. (1998). Situation models in language comprehension and memory. *Psychology Bulletin*, *123*, 162–185.

(RECEIVED 5/2/00; REVISION ACCEPTED 7/16/00)