Research Report

LANGUAGE COMPREHENDERS MENTALLY REPRESENT THE SHAPES OF OBJECTS

Rolf A. Zwaan, Robert A. Stanfield, and Richard H. Yaxley

Florida State University

Abstract—*We examined the prediction that people activate perceptual symbols during language comprehension. Subjects read sentences describing an animal or object in a certain location. The shape of the object or animal changed as a function of its location (e.g., eagle in the sky, eagle in a nest). However, this change was only implied by the sentences. After reading a sentence, subjects were presented with a line drawing of the object in question. They judged whether the object had been mentioned in the sentence (Experiment 1) or simply named the object (Experiment 2). In both cases, responses were faster when the pictured object's shape matched the shape implied by the sentence than when there was a mismatch. These results support the hypothesis that perceptual symbols are routinely activated in language comprehension.*

Consider the sentences *The ranger saw the eagle in the sky* and *The ranger saw the eagle in its nest*. According to most theories of language comprehension, the linguistic input would be converted to a propositional representation (e.g., Kintsch, 1998; Kintsch & van Dijk, 1978) such as [[SAW[RANGER,EAGLE]], [IN[EAGLE,SKY]]] and [[SAW[RANGER,EAGLE]], [IN[EAGLE,NEST]]]. Thus, the propositional representations for the two sentences would be largely identical, with the exception of the noun specifying the location. However, intuition suggests that this cannot be the whole story. After all, when a bird is in the air, it usually has its wings outstretched, and when it is in its nest, it usually has its wings folded. These differences are not captured in an amodal propositional structure like the one just given, although such a structure is routinely assumed by language comprehension researchers. Figure 1 illustrates a similar example. The shape of an egg is different when it is in the refrigerator than when it is in a skillet.

Inspired by a philosophical tradition that has thus far remained outside the mainstream of cognitive science, Barsalou (1999) recently argued that perceptual representations rather than amodal propositions are the building blocks of cognition. Perceptual symbols are the residues of a perceptual experience, stored as patterns of activation in the brain. Because attention is limited, perceptual symbols are typically schematic, rather than being akin to high-resolution video clips or high-fidelity sound clips. However, unlike amodal propositions, perceptual symbols bear an analog relationship with their referents. Barsalou hypothesized that perceptual symbols are used in perceptual simulations that make up human cognitive processes.

In a recent study (Stanfield & Zwaan, 2001), we found support for this idea in the domain of language comprehension. We presented subjects with sentences such as *He hammered the nail into the wall* or *He hammered the nail into the floor*. In the first sentence, the nail's orientation is horizontal, whereas in the second sentence it is vertical. Each sentence was followed by a line drawing of an object. For the experimental items, this was always the object whose implied orientation

was being manipulated in the sentence. The object in the drawing was presented either in a horizontal or in a vertical orientation, thus creating a match or a mismatch with the orientation implied by the sentence. The subjects made speeded recognition responses as to whether the object in the picture was mentioned in the sentence.

We tested two competing predictions. Perceptual symbol theories assume that people activate and manipulate perceptual symbols during language comprehension, such that an object's implied orientation in a sentence would be part of the mental representation of that sentence. Thus, according to such theories, responses would be faster when the object's implied orientation in the sentence matched the object's orientation in the picture compared with when there was a mismatch between the implied and pictured orientations. In contrast, in an amodal propositional representation the object's orientation would not be represented. Thus, according to amodal symbol theories, the match-mismatch manipulation would not affect response latencies to the picture. Our findings supported the perceptual symbol hypothesis. Responses were significantly faster when there was a match between implied orientation and pictured orientation than when there was a mismatch.

The purpose of the present study was to extend these findings in two ways. First, if language comprehenders represent the implied orientation of objects, they should also represent their implied shape. Thus, there should be a mismatch effect when subjects are presented with *The ranger saw the eagle in the sky* followed by a picture of an eagle with folded wings and also when subjects are presented with *The ranger saw the eagle in its nest* followed by a picture of an eagle with outstretched wings. We tested this hypothesis in two experiments. In Experiment 1, we used the same recognition paradigm as in our previous study (Stanfield & Zwaan, 2001). However, in a second extension of our earlier work, we used a naming task in Experiment 2. The naming task arguably provides a stronger test of the perceptual hypothesis in that unlike a recognition task it does not require an explicit comparison between the sentence and the picture.
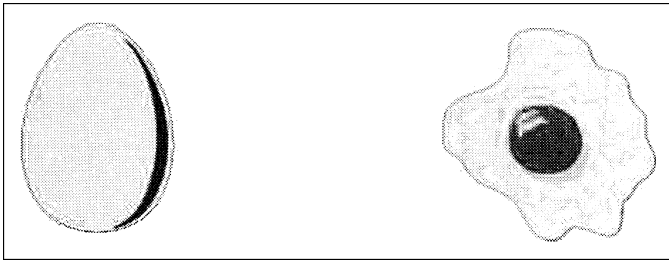
EXPERIMENT 1

Subjects

Fifty-one undergraduate students enrolled in introductory psychology courses at The Florida State University participated for course credit. The data of 7 subjects were discarded because of computer problems (not all the data for a given subject were recorded or the data file was corrupted). The data of 2 additional subjects were discarded because of extremely long median response latencies (>1,300 ms).

Materials

Seventy-two black-and-white drawings obtained from Snodgrass and Vanderwart (1980) and from a popular clip-art package were used. Of these pictures, 24 were used to construct filler items. The remaining 48 experimental pictures formed pairs, with the two members of a

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

168 VOL. 13, NO. 2, MARCH 2002

Rolf A. Zwaan, Robert A. Stanfield, and Richard H. Yaxley

Fig. 1. Different shapes of an egg: in a refrigerator versus in a skillet.

Table 1. *Object recognition latencies and accuracy in Experiment 1 and picture naming times in Experiment 2*

Measure	Condition		
	Match	Mismatch	Neutral
	Experiment 1		
Reaction time	697 (202)	761 (210)	—
Percentage correct	97 (6)	93 (7)	—
	Experiment 2		
Reaction time	605 (115)	638 (128)	617 (125)

Note. Standard deviations are given in parentheses.

pair showing different shapes of the same object. For example, one member of the pair might be a picture of an eagle with wings outstretched as if in flight and the other member a picture of an eagle with wings drawn in, as if perched. Other animals and objects used included an egg (in a carton vs. in a pan), an onion (in a basket vs. in batter), a frog (sitting vs. leaping), a book (on a table vs. on a photocopier), and bread (a loaf vs. a slice). Each picture was scaled to occupy a square of about 3 in.

Seventy-two sentences were created to accompany the pictures: 24 filler sentences and 48 experimental sentences. The experimental sentences were organized in pairs, with the two members of each pair implying different shapes of the same object. The filler sentences all mentioned an object (by way of a concrete noun) other than the one that was presented in the picture, and thus required a "no" response on the recognition task. The experiment was run on a PowerMac 7200/120 with an Apple Multiple Scan 15 Display using the Psyscope software program (Cohen, MacWhinney, Flatt, & Provost, 1993). Responses were recorded via the keyboard, using the "x" for "no" responses and the period key for "yes" responses.

Design and Procedure

We created four lists that counterbalanced items and conditions. Each list included a different one of the four possible versions (2 sentences × 2 pictures) for each object. Each subject saw one of these lists. This produced a 2 (condition: match vs. mismatch) × 2 (picture version) × 2 (list) design, with condition and shape (picture version) within-subjects variables and list a between-subjects variable. Thus, each subject saw 24 experimental sentence-picture pairs (12 match and 12 mismatch), requiring "yes" responses and 24 filler pairs, requiring "no" responses.

Subjects were instructed to read each sentence, and then to decide if the pictured object that followed had been mentioned in the preceding sentence. Subjects were further told that reaction times were being measured and that it was important for them to make the decisions about the pictures as quickly as possible. During each trial, subjects first saw a sentence, left-justified on the screen, that either mentioned or did not mention the object they would later see. They pressed the space bar when they had understood the sentence, and then a fixation point appeared in the center of the screen for 250 ms, followed by a picture. Subjects then determined if the pictured item had been mentioned in the previous sentence. The experiment took approximately 30 min to complete.

Results and Discussion

Table 1 displays the mean of the median response latencies as well as response accuracy for each condition. (Median response latencies

were used rather than means because of the within-subjects variability. However, analyses done on the averages yielded the same statistical pattern as the analyses with the medians.) We conducted a 2 (condition: match vs. mismatch) × 2 (picture version) × 2 (list) analysis of variance (ANOVA), with list as the only between-subjects variable, on the recognition response latencies and accuracy.

There was a significant mismatch effect on response latency: Responses were faster when sentence and picture matched than when they mismatched, $F_1(1, 38) = 13.14$, $p < .001$; $F_2(1, 44) = 14.54$, $p < .0001$. The two-way interaction between condition and list was not significant, $F_1(1, 38) = 3.55$, $p < .07$; $F_2 < 1$. The interaction between condition and picture version was significant in the analysis by items only, $F_1 < 1$; $F_2(1, 46) = 7.04$, $p < .015$. The three-way interaction involving all three factors was not significant, $F_1 < 1$; $F_2(1, 44) = 2.10$, $p > .15$.

Analyses of response accuracy showed that responses were more accurate when there was a match than when there was a mismatch, but this effect was significant in the analysis by subjects only, $F_1(1, 38) = 12.69$, $p < .001$; $F_2(1, 44) = 1.26$, $p > .25$. The Condition × List interaction was significant in the analysis by items only, $F_1(1, 38) = 1.20$, $p > .25$; $F_2(1, 44) = 9.05$, $p < .005$. The interaction between condition and picture version was not significant, $F_1(1, 38) = 1.47$, $p > .2$; $F_2(1, 44) = 1.75$, $p > .15$. The three-way interaction was not significant by subjects, but was significant by items, $F_1 < 1$; $F_2(1, 44) = 13.04$, $p < .001$.

These results support the prediction derived from perceptual symbol theory. Apparently, subjects represented the implied shape of the object when comprehending the sentence, so that responses to the picture were slower when the picture mismatched the implied shape than when there was a match between the pictured and implied shapes. The goal of Experiment 2 was to examine whether the same effect could be obtained with a task that does not call for a comparison between the picture and the sentence. In this experiment, the subjects merely named the picture after having read the sentence.

We also included a neutral condition in Experiment 2. The sentences in this condition did not imply anything about the shape of the object (e.g., *The ranger heard the eagle in the forest*). We included this condition to explore whether the mismatch effect observed in Experiment 1—and our previous study (Stanfield & Zwaan, 2001)—was due to a response facilitation in the match condition or a response inhibition in the mismatch condition. If the results were due to facilitation, response times in the neutral and mismatch conditions would be equal;

if the results were due to inhibition, response times in the neutral condition and match condition would be equal.

EXPERIMENT 2

Subjects

Fifty-seven undergraduate students enrolled in introductory psychology courses at The Florida State University participated for course credit. The data of 6 of those subjects were discarded because of the same computer problems as in Experiment 1.

Materials

The same line drawings and sentences as in Experiment 1 were used. An additional 24 neutral sentences that did not suggest a particular shape for the relevant object were created for the neutral condition. The experiment was run on a PowerMac 7200/120 with an Apple Multiple Scan 15 Display using Psyscope (Cohen et al., 1993). Responses were recorded using a Koss SB-30 headset-microphone attached to a Carnegie Mellon University button box.

Design and Procedure

We created six lists that counterbalanced items and conditions. Each list included a different one of the six possible versions (3 sentences $\times$ 2 pictures) for each object. Each subject saw one list. Thus, each subject saw 24 filler items, 8 match items, 8 mismatch items, and 8 neutral items. This design produced a 3 (condition: match vs. mismatch vs. neutral) $\times$ 2 (picture version) $\times$ 3 (list) design, with list as the only between-subjects variable. The procedure was nearly identical to that of Experiment 1. The only difference was that, instead of deciding whether the pictured object had been mentioned by the preceding sentence, subjects named the object.

Results

Table 1 shows the average naming time for each of the three conditions.[1] A 3 (condition: match vs. mismatch vs. neutral) $\times$ 2 (picture version) $\times$ 3 (list) ANOVA with list as the only between-subjects factor yielded a significant overall effect of condition by subjects, $F_1(2, 90) = 3.71$, $p < .04$; $F_2(2, 80) = 2.21$, $p < .12$. Picture version did not significantly interact with condition, $F_1(2, 90) = 1.78$; $F_2(2, 80) = 2.07$, $p > .13$, and neither did list, $F_1(4, 90) = 2.20$; $F_2 < 1$. The three-way interaction among condition, picture version, and list was significant in the analysis by items only, $F_1 < 1$; $F_2(4, 80) = 3.05$, $p < .025$.

Follow-up analyses comparing pairs of conditions showed that the mismatch condition yielded significantly slower responses than the match condition, $F_1(1, 45) = 6.90$, $p < .015$; $F_2(1, 40) = 3.87$, $p < .06$. No interactions were significant. The neutral condition was not significantly different from the match condition, $F_1 < 1$; $F_2(1, 40) = 2.86$, $p = .10$. No two-way interactions were significant (all Fs < 1). However, there was a significant three-way interaction involving condition, picture version, and list in the analysis by items only, $F_1 < 1$;

---

1. One item turned out to be problematic (an undeployed parachute, which could be mistaken for a backpack) and yielded unusually long naming times. It was omitted from the analyses.

$F_2(2, 40) = 4.64$, $p < .025$. The neutral condition was marginally significantly faster than the mismatch condition in the analysis by subjects only, $F_1(1, 45) = 4.01$, $p = .05$; $F_2 < 1$.

Discussion

These results extend those of Experiment 1 in that they show a mismatch effect even when the task, naming, does not call for a comparison between the sentence and the picture. It is interesting to note that the size of this effect was comparable across the two experiments: $d = .32$ in Experiment 1 and $d = .27$ in Experiment 2.

The results do not support either of the predictions pertaining to the neutral condition. It does not appear that the mismatch effect is due to a facilitation of responses in the match condition relative to the mismatch and neutral conditions, given that naming was not significantly slower in the neutral condition than in the match condition. Similarly, it does not appear that the mismatch effect is due to an inhibitory effect in the mismatch condition because naming was not significantly faster in the neutral condition than in the mismatch condition. However, the fact that the mean naming time in the neutral condition fell in between the naming times for the match and mismatch conditions is consistent with a scenario that can be derived from perceptual symbol theory. According to this scenario, comprehenders routinely activate perceptual symbols that include the shape of objects, even when the shape is not implied or articulated by the linguistic input. If this is the case, then given the counterbalancing scheme, a match between the sentence and the picture would be expected in half the observations and a mismatch would be expected in the other half. As a consequence, the average naming latencies in the neutral condition would fall between those of the match and mismatch conditions. This speculative scenario provides a post hoc explanation for the observed pattern of results. More extensive research is needed before firmer conclusions can be reached regarding the default representation of object shapes.

CONCLUSION

Our results, along with those of our previous study (Stanfield & Zwaan, 2001) and those of other recent studies (Fincher-Kiefer, in press; Kaschak & Glenberg, 2000; Kellenbach, Wijers, & Mulder, 2000), support the idea that people activate perceptual symbols of referents during language comprehension (Barsalou, 1999), even when the perceptual characteristics are merely implied rather than explicitly stated. Moreover, our results show that the sentential context has a strong and rather immediate impact on the nature of the mental representation (see also Hess, Foss, & Carroll, 1995; van Berkum, Hagoort, & Brown, 1999). These findings are consistent with the idea (e.g., Langacker, 1987) that the representation of meaning from linguistic input is a dynamic process involving malleable perceptual representations rather than the mechanical combination of discrete components of meaning (see also Barsalou, 1999; Glenberg, 1997; MacWhinney, 1999). The challenge for future research is to examine in more detail the theoretical and empirical ramifications of such a perceptual view of language comprehension.

Rolf A. Zwaan, Robert A. Stanfield, and Richard H. Yaxley

REFERENCES

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

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.

Fincher-Kiefer, R. (in press). Perceptual components of situation models. *Memory & Cognition*.

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

Hess, D.J., Foss, D.J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, *124*, 62–82.

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*, 508–529.

Kellenbach, M.L., Wijers, A.A., & Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: Event-related potential evidence for perceptual semantic priming. *Cognitive Brain Research*, *10*, 67–75.

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

Kintsch, W., & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Langacker, R.W. (1987). *Foundations of cognitive grammar* (Vol. 1). Stanford, CA: Stanford University Press.

MacWhinney, B. (1999). The emergence of language from embodiment. In B. MacWhinney (Ed.), *The emergence of language* (pp. 213–256). Mahwah, NJ: Erlbaum.

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., & Zwaan, R.A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*, 153–156.

van Berkum, J.J.A., Hagoort, P.M., & Brown, C.M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.