ELSEVIER

Cognitive Science 28 (2004) 611–619

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Moving words: dynamic representations in language comprehension☆

Rolf A. Zwaan*, Carol J. Madden, Richard H. Yaxley, Mark E. Aveyard

Department of Psychology, Florida State University, Tallahassee, FL 32306-1270, USA

Received 16 August 2003; received in revised form 3 March 2004; accepted 22 March 2004

Available online 18 May 2004

Abstract

Eighty-two participants listened to sentences and then judged whether two sequentially presented visual objects were the same. On critical trials, participants heard a sentence describe the motion of a ball toward or away from the observer (e.g., "The pitcher hurled the softball to you"). Seven hundred and fifty milliseconds after the offset of the sentence, a picture of an object was presented for 500 ms, followed by another picture. On critical trials, the two pictures depicted the kind of ball mentioned in the sentence. The second picture was displayed 175 ms after the first. Crucially, it was either slightly larger or smaller than the first picture, thus suggesting movement of the ball toward or away from the observer. Participants responded more quickly when the implied movement of the balls matched the movement described in the sentence. This result provides support for the view that language comprehension involves dynamic perceptual simulations.

Keywords: Visual motion; Implied motion; Perceptual representations; Language comprehension; Mental simulation; Embodied cognition

1. Introduction

In a famous experiment, Loftus and Palmer (1974) showed participants short traffic safety films of car accidents. They subsequently asked participants in different conditions how fast

☆ Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cogsci.2004.03.004.

* Corresponding author.

E-mail address: zwaan@psy.fsu.edu (R.A. Zwaan).

0364-0213/$ – see front matter
doi:10.1016/j.cogsci.2004.03.004

the cars were going when they 'contacted,' 'bumped into,' 'collided with,' 'smashed into,' or 'hit' each other. The speed implied by the verb in the question affected participants' judgments, such that more intense verbs led to higher estimates than less intense verbs. This finding has most often been used to demonstrate the unreliability of eyewitness testimony, or its malleability by post-event questions, but there is another intriguing aspect of this finding that has not received much attention. The finding suggests that language comprehension may affect the visual representation of a motion event. In this article, we examine whether language comprehension routinely involves the activation of visual motion representations. This question is warranted by findings in two domains: imagery and language comprehension. Specifically, there is empirical evidence that cognition involves dynamic visual representations, and there is evidence that language comprehension routinely involves the activation of (static) visual representations. These bodies of findings are discussed in the next two sections.

1.1. Dynamic representations

Research on what has become known as 'representational momentum,' has shown that mental representations can be dynamic (e.g., Freyd & Finke, 1984; Hubbard & Barucha, 1998; see Shepard & Metzler, 1971, for a seminal study on dynamic mental representations). For example, in the prototypical paradigm, participants are presented with a sequence of three pictures of a rectangle, with each rectangle slightly rotated relative to the previous one to suggest continuous rotation. Participants are then shown a recognition probe presented 250 ms after the last of the three-picture sequence. They generally false alarm to probes depicting the rectangle as slightly more rotated in the direction of the implied rotation than the last seen shape actually was (Freyd & Finke, 1984). Findings such as these are explained by assuming that the rectangle continued its rotation in the participants' mental representation after the last picture was shown, hence the term representational momentum. The extent to which representational momentum occurs appears to depend on context. For example, a sequence of three pointed shapes moving upward on a computer screen produces more representational momentum when the participant is told the shape represents a rocket ship than when the participant is told it represents a steeple (Reed & Vinson, 1996; Vinson & Reed, 2002). Brain-imaging studies have provided converging evidence for dynamic mental representations. Still pictures of animate entities in motion generate more activity in the medial temporal cortex, a brain region involved in motion perception, than do pictures of those same entities in a stationary position (Kourtzi & Kanwisher, 2000). There currently exists no unified theoretical explanation for findings such as these (Thornton & Hubbard, 2002), but the evidence is commonly interpreted as supporting the notion of dynamic mental representations (e.g., Freyd, 1987; Wallis & Bülthoff, 1999). In accordance with Wallis and Bülthoff, among others, we assume that dynamic mental object representations are the result of spatio-temporal associations between visual patterns acquired during experience of our environment.

1.2. Perceptual representations in language comprehension

The assumption that language-like abstract representations, such as propositions, form the building blocks of cognition (e.g., Pylyshyn, 1986) has recently been challenged by re-

search demonstrating that people routinely activate perceptual representations during language comprehension—the cognitive skill that intuitively would seem to be the one most likely to involve propositional representations—(Richardson, Spivey, Barsalou, & McRae, 2003; Spivey & Geng, 2001; Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002; Zwaan & Yaxley, 2003a, 2003b). Importantly, these studies employed implicit tasks, such as recognition or naming, thus demonstrating that perceptual information is routinely activated, even when doing so does not facilitate task performance.

These findings are consistent with the general idea that language comprehension is a perceptual simulation of the described situation (Barsalou, 1999; Glenberg, 1997; Zwaan, 2004). During our interaction with the world, we store traces in memory of perceptions and actions filtered through selective attention. These traces become associated with words (themselves traces of perceiving or producing sound or visual patterns). During language comprehension, these traces are reactivated to produce a perceptual simulation of the described situation. This line of reasoning provides a straightforward explanation for the findings of Zwaan et al. (2002). Participants in that study read sentences from a computer screen and were then shown a picture, which they had to recognize (Experiment 1) or name (Experiment 2). On experimental trials, the picture always showed an object or animal that was mentioned in the sentence. However, the shape of this entity was manipulated to match or mismatch the shape implied by the sentence. Thus, a picture of an eagle could follow a sentence such as “He saw an eagle in the sky” with its wings outstretched (match) or with its wings drawn in (mismatch). Both recognition and naming responses were significantly faster in the match than in the mismatch condition. This finding is not predicted by amodal theories of cognition (e.g., Pylyshyn, 1986). These theories represent the eagle as an argument node in a propositional network, or as a list of features, thus failing to capture the fact that its shape may change according to the location that it is in. In contrast, the perceptual-simulation hypothesis has a natural account for this finding. Visual traces of soaring and perched eagles are stored in memory. These traces are reactivated during comprehension, with the most contextually consistent traces receiving the most activation and thus being the most likely to be incorporated in the simulation. Seeing the picture during the experiment produces a new visual trace. In the match condition, this trace will be more similar to the trace activated by the sentence than in the mismatch condition, so that a comparison can be made more rapidly, thus producing faster recognition and naming responses.

2. The present study

In this study, we combined and extended the logic behind theories of representational momentum and theories of language comprehension as perceptual simulation. We assume that dynamic mental representations are perceptual traces that are stored as temporal patterns of activation that unfold over time corresponding to a certain perceptual experience. Extending the logic behind the Stanfield and Zwaan (2001) and the Zwaan et al. (2002) experiments, we predicted that comprehension of a sentence describing a motion event should facilitate the perception of an analogous visual motion event relative to a mismatching motion event. We tested this idea using sentences such as (1) and (2).

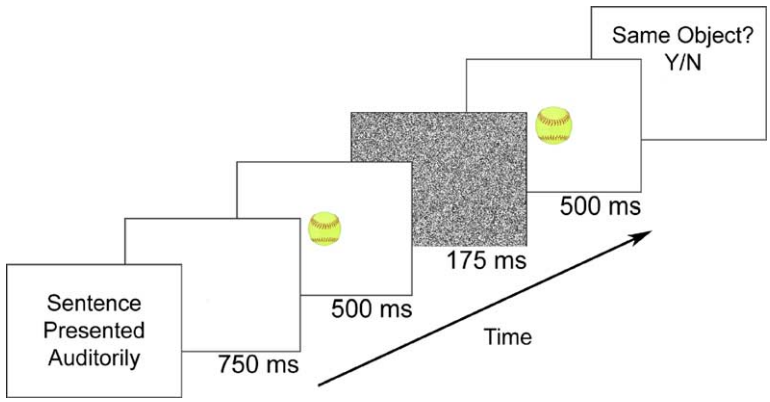

Fig. 1. Event sequence on a given trial.

(1) The shortstop hurled the softball at you.
(2) You hurled the softball at the shortstop.

After each—auditorily presented—sentence, a sequence of two pictures was shown. On critical trials, two pictures of the same object were each shown for 500 ms, separated by a 175 ms mask, with the first object being slightly larger or smaller than the second one. A bigger object following a smaller one would suggest movement toward the observer, whereas a smaller object following a bigger one would suggest movement away from the observer. A sentence such as (1) followed by a smaller–bigger sequence would produce a match, whereas the same sequence would produce a mismatch for sentence (2). The reverse is true for a bigger–smaller sequence. Fig. 1 displays the sequence of events during a given trial.

The participants indicated whether the two pictures displayed the same object. In order to make this a very easy perceptual task, we included filler trials in which we presented two entirely different pictures. In order to ensure that the participants were not ignoring the sentences, we asked comprehension questions after a subset of our sentences. Our prediction, based on the notion of perceptual simulation, was that participants should be faster at judging pictures in the match condition than in the mismatch condition, despite the fact that the comprehension task is not relevant to the perceptual task.

2.1. Method

2.1.1. Participants

Eighty-two students enrolled at Florida State University participated in the experiment as part of a course requirement. All participants were native English speakers.

2.1.2. Materials

Eighty-six sentences were constructed. These sentences were spoken by a male native speaker of American English and digitally recorded on a PC with a professional Aardvark© Q10 sound card and pre-amplifier using a Shure© SM58 cardioid dynamic microphone. The

sound files were digitally equalized, compressed, and edited using Cakewalk© Sonar© XL 2.0 with Soundforge© 5.0 plug-ins.

Eighty-six color pictures were constructed to accompany the sentences. All pictured objects were scaled to three size levels (7, 8, and 9 cm). Twenty of these pictures were experimental pairs. All 20 critical trials consisted only of round balls (e.g., basketball, beach ball, ping pong ball) in order to maintain radial symmetry between the toward and away conditions. Sixty-six pictures were used as fillers (six as practice trials). Eleven round balls were included in the filler items in order to disguise the critical trials. Implied motion was varied within the pictures to match/mismatch the direction of motion described in the experimental sentences. Implied motion towards the observer was depicted by presenting a small picture followed by a medium picture, whereas implied motion away from the observer was depicted by presenting a large picture followed by a medium picture; thus, responses were always made to the medium-sized picture. The actual size difference between the medium-sized items and the large and small ones was 0.9 cm for the longest dimension. Specifically, the maximum width or height for our large pictures was 8.8 cm, whereas it was 7 cm for the small pictures, and 7.9 cm for the medium pictures. In other words, the size manipulation was quite subtle. A pair of sentence examples for the away and toward conditions are: "You tossed the beach ball over the sand toward the kids," and "The kids tossed the beach ball over the sand toward you." The complete set of stimulus items is available at http://cognitivesciencesociety.org/supplements/.

2.1.3. Procedure

Four lists were constructed to counterbalance items and conditions. Each list included one of four possible versions: 2 (sentence: toward/away) × 2 (picture: bigger/smaller) for each object. Each participant saw only one list. This produced a 2 (sentence) × 2 (picture) × 4 (list) design. Sentence and picture were within-participants and item variables, and list was a between-participants variable. Match and mismatch conditions were balanced across all lists. Each participant saw 20 experimental sentence–picture pairs (10 match and 10 mismatch), requiring 'yes' responses, 20 filler pairs requiring 'yes' responses, and 40 filler pairs requiring 'no' responses. Thus, there were 40 sentence–picture pairs requiring 'yes' responses and 40 requiring 'no' responses. There were six practice items, three with comprehension questions.

The experiment was run on PCs with 19 in. flat-screen displays using the E-Prime stimulus presentation software (Schneider, Eschman, & Zuccolotto, 2002). Each trial consisted of the following sequence of events: (1) a sentence presented auditorily over headphones, (2) the first picture (big or small) presented for 500 ms, (3) a full-screen black-and-white random-dissolve patterned mask presented for 175 ms, and (4) a second (medium) picture presented for 500 ms. Participants were instructed to listen to the sentences and then judge if the two pictures displayed the same object. Furthermore, the participants were instructed to respond quickly and accurately as both reaction time and accuracy of response were being measured. Responses were recorded via the keyboard using the 'Y'-labeled J-key for 'yes' responses and the 'N'-labeled F-key for 'no' responses. Because the picture task could be performed without attending to the sentences, comprehension questions were presented after 16 of the filler trials to ensure that participants were paying attention to the sentences (e.g., "Were you worried about your knee injury?"). Participants were instructed to use the yes and no keys to answer these comprehension questions.

Table 1
Average reaction times by condition (accuracy in parentheses)

Toward pictures		Away pictures	
Match (toward sentences)	499 (0.99)	Match (away sentences)	492 (0.98)
Mismatch (away sentences)	511 (0.98)	Mismatch (toward sentences)	522 (0.96)

2.2. Results

Responses greater than 1500 ms as well as responses that were above or below two standard deviations from a participant's condition mean for correct RTs were excluded. Overall, this led to the exclusion of less than 2% of the data. Table 1 shows the average response times and accuracy, segregated by type of picture sequence (toward-away), match or mismatch, and type of sentence (toward-away).

Mixed analyses of variance (ANOVAs) were conducted on the response times with match and direction (toward or away from the protagonist) as within-participants variables and list as a between-participants variable.[1] Effects involving list will not be reported because of their theoretical irrelevance (Pollatsek & Well, 1995). The relevant means and standard errors are displayed in Table 1. Responses were significantly faster when the direction implied by the visual presentations matched that of the sentence than when there was a mismatch ($F_1(1, 77) = 6.83$, $p < .015$, $MS_e = 5177$; $F_2(1, 16) = 38.04$, $p < .0001$). Furthermore, participants did not respond more quickly when the sentences described the ball as moving away from than toward the observer (both Fs < 1). The interaction between these factors was not significant by participants, but did reach significance by items ($F_1(1, 77) = 2.74$, $p > .10$; $F_2(1, 16) = 5.50$, $p < .05$). This pattern is due to the fact that the mismatch effect was larger for the "away" picture sequences (30 ms) than for the "towards" sequences (12 ms). Because this interaction did not reach conventional levels of significance in the analysis by participants, we will not interpret it further. Finally, there were no significant effects regarding response accuracy (all ps $> .10$).

2.3. Discussion

As predicted by the perceptual-simulation hypothesis, responses were faster when the picture sequence matched the movement of the ball as described by the sentence. For example, if the sentence described the ball as moving away from the protagonist, then participants responded more quickly when a larger picture of a ball preceded a medium-sized picture of that same ball (suggesting movement away) than when a smaller picture preceded the medium-sized picture. The reverse was true for sentences implying movement toward the protagonist. It is important to note that this effect of sentence content occurred in a task in which participants made speeded decisions as to whether two successively presented pictures depicted the same object, a task for which the sentence content was irrelevant.

The effect can be explained by assuming that sentence comprehension involves the perceptual simulation of the events described in a sentence. The mental representation generated by viewing the picture sequence following the sentence either matched that simulation or mis-

matched it, leading to faster responses in case of a match than in case of a mismatch. "Match" and "mismatch" should be interpreted in relative terms here. We do not claim that the memory representation produced during sentence comprehension is identical to the one generated during picture viewing. Our claim is simply that the match between these two representations is greater in the match condition than in the mismatch condition, such that the match condition receives more priming from the sentence-induced perceptual simulation than the mismatch condition.

3. Conclusion

The findings reported in this article add to the growing body of evidence that language comprehension routinely involves the activation of perceptual representations (Pecher, Zeelenberg, & Barsalou, 2003; Richardson et al., 2003; Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan & Yaxley, 2003a, 2003b). However, the current findings constitute an advance over this earlier research in several ways. Most importantly, they demonstrate that language comprehension may involve dynamic mental representation. During our interactions with objects, we acquire dynamic mental representations of their movements. Thus, our dynamic visual representation of a ball moving toward us involves that of an object rapidly occupying more of our visual field. Conversely, our dynamic representation of a ball moving away from us involves that of an object rapidly occupying less of our visual field. We were able to approximate this experience by presenting in quick succession two pictures of the same object with the first one being either slightly smaller or larger than the second one.

This study also makes a novel contribution to the study of dynamic mental representations. The occurrence of this type of representation is typically demonstrated by having participants view a sequence of pictures implying a certain type of motion (e.g., rotation around an object's axis), followed by a probe, which is a picture of the same object, either further along its path of rotation or shown in the same orientation as the last picture. Thus, in this type of experiment, the representational motion is generated by an actual visual display. In our experiment, we argue, the motion is generated entirely by a linguistic stimulus, a sentence describing motion. One earlier study (Reed & Vinson, 1996; see also Vinson & Reed, 2002) has used verbal stimuli to manipulate dynamic mental representations. A sequence of pictures was shown implying upward movement. The pictures were abstract enough such that they could be interpreted as a rocket ship or a steeple. Participants demonstrated more representational momentum when they interpreted the stimulus as a rocket ship than when they interpreted the stimulus as a steeple. There are two main differences between this study and our experiment. First, in the Reed and Vinson study, representational motion was created by a sequence of pictures and it was the degree of representational motion that was modulated by the verbal labels. In our experiment, representational motion was generated solely by the sentences. Second, in the Reed and Vinson study, participants were explicitly told to interpret the visual stimuli as the object denoted by the label. In our experiment, there was no direct connection between the visual stimuli and the sentence, and performance on the picture comparison task could strictly speaking be performed independently of the sentences. As such, our procedure can be considered a more implicit manipulation of the effect of linguistic context on dynamic mental representations. Third, our

verbal stimuli were sentences, rather than individual words and because participants were not told to identify the pictures with the sentences, our experiment can be said to provide a more naturalistic test of the effect of language comprehension on representational motion. As such our results both support and extend earlier demonstrations of dynamic mental representations and thus prompt further theorizing on this topic.

To complete the circle started in our introduction, our findings provide an explanation for the well-known findings of Loftus and Palmer (1974). The verbs used to probe the participants' memory for the target event were cues to start dynamic perceptual simulations. Verbs associated with greater speed will produce faster perceptual simulations (i.e., more perceptual change per time unit) than verbs associated with lower speeds. These simulations, rather than the initial memories, were then used to estimate the speed of the vehicle, suggesting that words can, indeed, move mental representations.

Note

1. It would have also been possible to use sentence type instead of match as a factor. In that case, an interaction between sentence type and picture type would be predicted. This analysis is essentially equivalent as the one reported here and so yields the same results. We chose to use match as a factor rather than sentence type for ease of exposition.

Acknowledgments

We thank Meredith Lynam, Michelle Peruche, Raymond Britton, and Greg Smith for assistance with data collection. This research was supported by grant MH-63972 to R.A.Z.

References

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

Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, *94*, 427–438.

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 126–132.

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

Hubbard, T. L., & Barucha, J. J. (1998). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, *44*, 211–221.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, *12*, 48–55.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning & Verbal Behavior*, *13*, 585–589.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychological Science*, *14*, 119–124.

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 & Cognition*, *21*, 785–794.

Pylyshyn, Z. W. (1986). *Computational cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

Reed, C. L., & Vinson, N. G. (1996). Conceptual effects on representational momentum. *Journal of Experimental Psychology: Human Perception & Performance*, *22*, 839–850.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*, 767–780.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime 1.0*. Pittsburgh, PA: Psychological Software Tools.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Spivey, M., & Geng, J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*, 235–241.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *121*, 153–156.

Thornton, I. M., & Hubbard, T. L. (Eds.). (2002). *Representational momentum: New findings, new directions*. New York: Psychology Press/Taylor & Francis.

Vinson, N. G., & Reed, C. L. (2002). Sources of object-specific effects in representational momentum. *Visual Cognition*, *9*, 41–65.

Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences*, *3*, 22–31.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35–62). New York: Academic Press.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science*, *13*, 168–171.

Zwaan, R. A., & Yaxley, R. H. (2003a). Hemispheric differences in semantic-relatedness judgments. *Cognition*, *87*, B79–B86.

Zwaan, R. A., & Yaxley, R. H. (2003b). Spatial iconicity affects semantic-relatedness judgments. *Psychonomic Bulletin & Review*, *10*, 954–958.