ARTICLE IN PRESS

Available online at www.sciencedirect.com
ScienceDirect

ELSEVIER

Cognition xxx (2006) xxx–xxx

COGNITION

www.elsevier.com/locate/COGNIT

Brief article

Simulating visibility during language comprehension

Richard H. Yaxley *, Rolf A. Zwaan

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

Received 20 July 2006; revised 11 August 2006; accepted 4 September 2006

Abstract

In this study, participants performed a sentence–picture verification task in which they read sentences about an agent viewing an object (e.g., moose) through a differentially occlusive medium (e.g., clean vs. fogged goggles), and then verified whether a subsequently pictured object was mentioned in the previous sentence. Picture verification latencies were shorter when the resolution of the pictured object and the resolution implied by the sentence matched than when they did not. These results suggest that the degree of visibility implied in linguistic context can influence immediate object interpretation. These data suggest that readers mentally simulate the visibility of objects during language comprehension. Thus, the simulation of linguistic descriptions is not limited to the activation of intrinsic object properties (e.g., object shape), but also invokes the perceptibility of referential objects given implied environmental context.
© 2006 Published by Elsevier B.V.

Keywords: Language comprehension; Mental simulation; Perceptual representations; Object visibility

1. Introduction

To what extent are described events processed like direct perceptual experiences? Understanding described events is often accompanied with a rich, and seemingly

* Corresponding author. Tel.: +1 850 644 1425.
E-mail address: yaxley@psy.fsu.edu (R.H. Yaxley).

0010-0277/$ - see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.cognition.2006.09.003

Please cite this article in press as: Yaxley, R. H., & Zwaan, R., Simulating visibility during language comprehension, *Cognition* (2006), doi:10.1016/j.cognition.2006.09.003

palpable, sense of 'being there'. A growing body of empirical demonstrations supports this notion. For example, several studies have recently demonstrated that language processing invokes perceptual and motoric activation that resembles the patterned activation involved in direct experience (Glenberg & Kaschak, 2002; Kaschak et al., 2005; Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002; Zwaan & Yaxley, 2003; Zwaan & Taylor, 2006). This select activation is indicative of a mental simulation that occurs when linguistic cues reactivate the experiential traces associated with described referents (Barsalou, 1999; Barsalou, Simmons, Barbey, & Wilson, 2003; Goldstone & Barsalou, 1998; Zwaan, 2004). Thus, the same cognitive mechanisms that are involved in the perception of, and action with, objects and events are also recruited, at least partially, in the mental representations of those objects and events. This process of mental simulation may be the keystone for grounding cognition meaningfully within environmentally situated agents (Glenberg, 1997; Harnad, 1990; for a review see Wilson, 2002). However, much remains unknown about how comprehension is influenced by the intricacies of mental simulations that develop as language unfolds (Zwaan, 2004). For instance, it is unknown if language-based mental simulations are driven solely by the activation of object properties, or if they also invoke perceptual traces associated with the qualitative experience of events in various situations (e.g., the visibility of objects). The present study was designed to investigate this question by testing if language comprehension activates the visibility of objects given environmental context.

Recent work has demonstrated that linguistic context does activate situation-specific object information (Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan & Yaxley, 2003). For example, Stanfield and Zwaan (2001) demonstrated that object orientation was activated during sentence comprehension. In this study participants read sentences such as 'The carpenter hammered the nail into the {floor/wall}', and subsequently verified whether a pictured object was mentioned in the sentence. Critically, the pictured object either matched or mismatched the implied orientation of the object (vertically oriented for …in the floor; horizontally oriented for …in the wall). Participants' verification times were longer when the orientation of the pictured objects was incompatible with the described events (e.g., a horizontally oriented nail after reading that the nail was hammered into the floor) than when the depicted and implied orientations were compatible. Using a similar sentence–picture verification task, Zwaan et al. (2002) demonstrated that object shape was also activated during language comprehension. Participants read sentences such as 'The park ranger saw the eagle in the {sky/nest}'. Then they saw a pictured object whose shape either matched the implied shape of the object (wings outstretched for …in the sky; wings withdrawn for …in the nest) or not. As in Stanfield and Zwaan (2001), participants' verification times were longer when the shape of the pictured objects was incompatible with the described events. These findings demonstrate that comprehenders activate contextually-appropriate perceptual information about described objects. However, the activation of perceptual content is not limited to the intrinsic attributes of described objects, because environmental context can also affect relevant perceptual information.

Interacting with the visual world involves the ongoing acquisition of object attributes (e.g., location, identity) amid the influence of the environmental context. For example,

Please cite this article in press as: Yaxley, R. H., & Zwaan, R., Simulating visibility during language comprehension, *Cognition* (2006), doi:10.1016/j.cognition.2006.09.003

imagine driving through a thick blanket of morning fog, only to see the shape of a dog emerge along the roadside. In this example, the availability of the resulting object representation (i.e., dog) should be modulated by the degree of occlusion (i.e., fog) present in the percept. If language processing recruits the same mechanisms that are involved in perception, then it stands to reason that described occlusion should also modulate the availability of the resulting representation. Horton and Rapp (2003) demonstrated that described occlusion influences the availability of objects during recall. In this study, participants read brief stories that described a protagonist's view of an object that was occluded by an intervening medium (e.g., fog, smoke, window blinds, etc.). Subsequently, participants were presented with the name of a target object and had to verify the presence of the target object in the story. Objects that were described as being blocked from view elicited slower verification judgments than for objects that remained in clear view. Thus, retrieval of a described object appears to be modulated by described environmental occlusion. This study raises an important question: do the representations underlying comprehension depend on intrinsic properties of referential objects or do they also invoke properties associated with implied environmental context.

In the current study, we investigated whether language processing invokes information about the visibility of objects by determining if described environmental situations influence immediate object verification.

Consider the following sentences:

(1) Through the fogged goggles, the skier could hardly identify the moose.
(2) Through the clean goggles, the skier could easily identify the moose.

Sentences (1) and (2) describe the same basic situation of an agent observing an object with a fundamental difference: the object is differentially (hardly vs. easily) visible by virtue of two different adjectives in the prepositional phrase (fogged vs. clean). Thus, the agent's observation occurs amid varying degrees of environmental occlusion. If language processing involves the contextually constrained activation of experiential traces, then the representations born from these two descriptions should differ across visual resolution. In other words, reading about a foggy moose should yield a mental representation akin to one that is evoked when actually seeing a foggy moose, and thus influence the actual perception of one. Therefore, comprehenders should be faster to recognize a picture of the described object that matches the level of visual resolution (i.e., high vs. low) than one that does not match. This study was designed to test if the level of visual resolution represented during language processing influences picture verification.

2. Present experiment

2.1. Method

2.1.1. Participants

Eighty undergraduate students enrolled in introductory psychology courses at Florida State University participated for course credit.

Please cite this article in press as: Yaxley, R. H., & Zwaan, R., Simulating visibility during language comprehension, *Cognition* (2006), doi:10.1016/j.cognition.2006.09.003

2.1.2. Materials

Ninety-six sentences were created: 24 experimental sentence pairs and 48 filler sentences. The experimental sentence pairs were of the form Through the 'medium X', the agent could see the 'object Y' [e.g., Through the {fogged/clean} goggles, the skier could {hardly/easily} identify the moose]. Thus, both members of each sentence pair implied varying degrees of object visibility. For example, 'Through the fogged goggles' implies that the visibility of the concrete noun (i.e., moose) is low, whereas, the 'Through the clean goggles' sentence implies that the visibility of the concrete noun is high. Twelve of the 48 filler sentences were of the same format as the experimental sentences in that they included an agent actively viewing a concrete object. The remaining 36 sentences included an agent interacting with a concrete object in a way that did not highlight the act of seeing. Twelve of the filler sentences were followed by a pictured object mentioned in the sentence and required 'yes' responses, and the 36 remaining filler sentences were followed by a pictured object not mentioned in the sentence and required 'no' responses.

Ninety-six black-and-white pictures were created. All pictured objects were scaled to approximately three inches. Twenty-four of these pictures were experimental pairs, with both members of each pair depicting the same object at two levels of spatial resolution. Forty-eight pictures were used as fillers. Spatial resolution of the pictures was varied to match/mismatch the resolution level implied by the occlusive medium described in the experimental sentences (e.g., fog, smoke, mist, etc.). Thus, one member of the experimental pair was a clear picture of a moose as viewed through 'clean goggles' while the other was an unclear picture of the same moose as viewed through 'fogged goggles'. Spatial resolution was also varied for the filler pictures so that there were an equal number of clear and unclear pictures across the experiment (see Fig. 1).

Two levels of object-resolution (clear and unclear) were varied among the pictures. Unclear images were generated with a software-based random-dissolve filter. The random-dissolve filter created pictures that were similar to what one

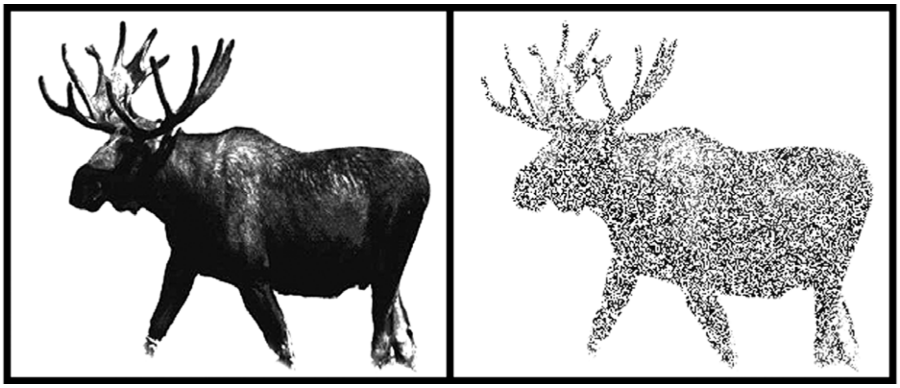

Fig. 1. Example high-resolution stimulus (left) and low-resolution stimulus (right).

Please cite this article in press as: Yaxley, R. H., & Zwaan, R., Simulating visibility during language comprehension, *Cognition* (2006), doi:10.1016/j.cognition.2006.09.003

might observe through a snowfall or a 'snowy' picture on a television, and in this sense the style of degradation was somewhat naturalistic. Extensive pilot testing revealed that participants were quite adept at identifying the pictured objects when the occlusion level was less than 50%. At 50% occlusion, however, there was an effect of picture on object recognition latencies (clear vs. unclear, $t(19) = -2.95$, $p < .01$, $MSe = 19.09$), such that clear pictures were responded to more quickly than unclear pictures. However, there was no speed-accuracy tradeoff between the pictures ($t(19) = .81$, $p < .44$), and the group mean accuracy scores remained greater than 95% for both types of pictures. Thus, the random-dissolve filter at 50% occlusion level ensured that our pictures were substantially degraded, but not excessively so, in that it yielded a difference in response latencies while not influencing accuracy scores.

2.1.3. Design and procedure

Four lists of stimuli were created to counterbalance items and conditions. Each participant saw one of four possible versions 2 (sentence: clear/unclear) × 2 (picture: clear/unclear) for each object. This produced a 2 (sentence) × 2 (picture) × 4 (list) design. Sentences and pictures were within-participants and within-item factors, and list was a between-participants factor. Match and mismatch conditions were balanced across all lists. Each participant saw 24 experimental sentence–picture pairs (12 match and 12 mismatch), requiring 'yes'; responses, 12 filler pairs requiring 'yes' responses, and 36 filler pairs requiring 'no' responses. Thus, there were 36 sentence–picture pairs requiring 'yes' and 36 requiring 'no' responses.

Stimulus presentation was controlled by the E-Prime software package (Schneider, Eschman, & Zuccolotto, 2002) on 19 in. flat-screen CRT displays. During each trial, a sentence was presented in the center of the screen and participants were instructed to read the sentence, to press the spacebar when finished reading, and then to judge whether the subsequently pictured object had been mentioned in the sentence or not. Participants were encouraged to respond quickly and accurately as both reaction times and response accuracy would be recorded. Responses were recorded via the keyboard using a 'Y'-labeled button for 'yes' responses and a 'N'-labeled key for 'no' responses.

3. Results and discussion

Response latencies and error rates of experimental trials were submitted to 2 (sentence) × 2 (picture) × 4 (list) analyses of variance (ANOVAs) with repeated measurement on pictures and sentences in both the by-participant and by-items analyses. All latency analyses were performed on correct responses only. Responses shorter than 300 ms and longer than 2000 ms were omitted, as well as those responses falling outside 2 standard deviations from the mean in each condition for each participant (this eliminated less than 2% of the data). The mean latencies, standard deviations, and standard errors are displayed in Table 1.

Table 1
Mean latencies/standard deviations of correct responses (in ms) and accuracy scores (percent correct) in the Picture Recognition Task

	Clear sentence *M*/SD	Unclear sentence *M*/SD
Clear picture	635/150	658/163
Unclear picture	699/154	670/147
	Accuracy	Accuracy
Clear picture	97%	98%
Unclear picture	98%	99%

Most critical to our predictions was a significant crossover interaction between pictures and sentences ($F_1(1,76) = 11.88$, $p < .001$, MSe = 55020; $F_2(1,20) = 14.74$, $p < .001$, MSe = 25513). Match between sentences and pictures elicited shorter picture-verification latencies than did mismatching pictures and sentences. We segregated the items by pictures to investigate this interaction further. Clear pictures were responded to more quickly when preceded by a clear sentence than when preceded by an unclear sentence ($F_1(1,76) = 4.36$, $p < .04$, MSe = 21739; $F_2(1,20) = 4.69$, $p < .05$, MSe = 12160), and unclear pictures were responded to more quickly when preceded by a unclear sentence than when preceded by a clear sentence ($F_1(1,76) = 6.44$, $p < .02$, MSe = 33960; $F_2(1,20) = 5.05$, $p < .04$, MSe = 13367). There was also a significant main effect of picture type (clear vs. unclear, $F_1(1,76) = 15.41$, $p < .001$, MSe = 106872; $F_2(1,20) = 5.51$, $p < .03$, MSe = 29786). This effect is not surprising given the results of the pilot study in which clear pictures were responded to more quickly than unclear pictures. These findings demonstrate that language processing invokes the visibility of objects given described environmental context. These findings converge with those of Horton and Rapp (2003) with a more implicit measure that takes into account how the perceptibility associated with described events influences perception.

4. Conclusion

This study was conducted to investigate whether comprehenders mentally simulate the implied visibility of objects. After reading a sentence that described a protagonist viewing a target entity through an obscuring medium, participants were faster to recognize a picture of the target entity if the depicted visual resolution matched the resolution implied in the sentence. These findings support the hypothesis that comprehenders represent the perceptibility of objects given described environmental context, and thus the prediction that language processing invokes experiential traces in a mental simulation of described events (Barsalou, 1999; Zwaan, 2004). Furthermore, the results provide converging evidence that language comprehension routinely activates higher-level (e.g., object accessibility during recall; Horton & Rapp, 2003) as well as lower-level (e.g., priming of object features like orientation and shape; Stanfield & Zwaan, 2001; Zwaan et al., 2002) semantic information. However, the current study advances beyond the previous findings by demonstrating that language comprehension utilizes implied situational context that modulates the perceptibility of a depicted referent. Thus, the modality-specific information activated during language comprehension affects perception in a qualitative manner similar to how comprehension affects overt action (e.g., as in Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006).

Converging evidence from a diverse set of paradigms suggests that sensorimotor representations are routinely activated during comprehension. The role that these sensorimotor representations play in supporting comprehension remains an open question. For instance, it is unclear if sensorimotor representations are an integral component of conceptual simulations or if they play an ancillary role in understanding. Thus, it is unknown to what extent environmental context and task-requirements influence the activation sensorimotor representations, and if comprehenders develop unique task-based strategies that facilitate or inhibit the activation of select experiential knowledge to maximize successful task performance. Although the strategical mechanisms that would be implemented in these aspects of conceptual processing remain elusive, [1] future work will need to investigate how dependent conceptual processing is to environmental context and task-performance.

The interactive influence between language and perceptual processing suggests that conceptual representations and the sensorimotor representations that give rise to them are from a unified representational system (Barsalou, 1999). We conclude from the findings obtained here that the representations underlying comprehension are qualitative analogs of patterned activation that are involved in perception and action. Furthermore, not only are such representations modality-specific in how they interactively influence perception (Pecher, Zanolie, & Zeelenberg, in press), but they also involve dynamic activation (Zwaan, Madden, Yaxley, & Aveyard, 2004). The present study also demonstrates that representations of objects can be integrated with representations of described environmental context to generate unique

[1] A reviewer suggested that the subjects might have employed some kind of verbalizing strategy in which they constructed a verbal description of the depicted entities (e.g., 'fogged moose'), and then compared this with the textbase associated with the preceding sentence (Kintsch, 1998). In this scenario, the priming effect would be due to lexical or propositional associations rather than to priming between visual representations. Conceptually, this is a reasonable characterization of what could be involved in the decision-making process, however it involves several, potentially unnecessary, assumptions about the processes underlying comprehension. First, adopting such a multi-dimensional verbalization strategy in the context of the present study seems less than optimal. In half of the trials, the pictured entity was not even mentioned in the previous sentence. Only half of the experimental trials involved a match between picture and sentence along both dimensions, and even fewer trials (<17% of total) were presented in which such a strategy would be useful. Second, employing such a strategy would increase the participants' workload to a level that is unnecessary for successful task completion. After all, participants were only instructed to verify whether the pictured object was mentioned in the previous step; it would be unnecessary to verbalize the adjective (e.g., "foggy") as well. Third, given that there were a variety of adjectives used across trials, participants would be responsible for constructing specific verbal descriptions for each of the pictures to successfully benefit from a two-dimensional match between textbase and picture verbalization. Fourth, the timescale of the verification latencies leaves scant time for the conscious construction of verbal descriptions along multiple dimensions. This seems especially true when one factors in the amount time necessary for perception and motor-planning. For these reasons, a strategy based on generating verbal descriptions seems an unlikely approach that would be adopted by participants.

Please cite this article in press as: Yaxley, R. H., & Zwaan, R., Simulating visibility during language comprehension, *Cognition* (2006), doi:10.1016/j.cognition.2006.09.003

conceptual representations. The fact that described environmental context of vicarious situations influences picture verification performance is indicative of a system that automatically recruits a rich set of experiential traces to support understanding despite the fact that the task did require such specific or exhaustive activation. The present findings provide strong support that the sensorimotor simulations involved in conceptual processing are qualitative analogs of modal experience that show an immediate influence the interpretation of the environment.

References

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

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences, 7*, 84–91.

Glenberg, A. M. (1997). Mental models, space, and embodied cognition. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 495–522). Washington, DC: American Psychological Association.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558–565.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition, 65*, 231–262.

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

Horton, W. S., & Rapp, D. N. (2003). Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review, 10*(1), 104–109.

Kaschak, M. P., Madden, C. J., Therrriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., et al. (2005). Perception of motion affects language processing. *Cognition, 94*(3), B79–B89.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge Unversity Press.

Pecher, D., Zanolie, K., & Zeelenberg, R. (in press). Verifying visual properties in sentence verification facilitates picture recognition memory. *Experimental psychology*.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools, Inc.

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

Wilson, M. (2002). Six view of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625–636.

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. 43, pp. 35–62). New York: Academic Press.

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

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology General, 135*, 1–11.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: dynamic mental representations in language comprehension. *Cognitive Science, 28*, 611–619.