Available online at www.sciencedirect.com
SCIENCE DIRECT
COGNITION
Cognition 94 (2005) B79–B89
www.elsevier.com/locate/COGNIT

Brief article

Perception of motion affects language processing

Michael P. Kaschak*, Carol J. Madden, David J. Therriault, Richard H. Yaxley, Mark Aveyard, Adrienne A. Blanchard, Rolf A. Zwaan

Department of Psychology, Florida State University, Tallahassee, FL 32310, USA

Received 20 February 2004; accepted 2 June 2004

Abstract

Recently developed accounts of language comprehension propose that sentences are understood by constructing a perceptual simulation of the events being described. These simulations involve the re-activation of patterns of brain activation that were formed during the comprehender's interaction with the world. In two experiments we explored the specificity of the processing mechanisms required to construct simulations during language comprehension. Participants listened to (and made judgments on) sentences that described motion in a particular direction (e.g. "The car approached you"). They simultaneously viewed dynamic black-and-white stimuli that produced the perception of movement in the same direction as the action specified in the sentence (i.e. towards you) or in the opposite direction as the action specified in the sentence (i.e. away from you). Responses were faster to sentences presented concurrently with a visual stimulus depicting motion in the opposite direction as the action described in the sentence. This suggests that the processing mechanisms recruited to construct simulations during language comprehension are also used during visual perception, and that these mechanisms can be quite specific.

Keywords: Language comprehension; Sentence comprehension motion perception; Embodied cognition

Recent theories of sentence processing suggest that sentence comprehension involves constructing a sensorimotor simulation of the described events (e.g. Glenberg & Kaschak, 2002; Kaschak & Glenberg, 2000; Stanfield & Zwaan, 2001; Zwaan, 2004). These simulations involve the re-activation of patterns of brain activation that were formed

* Corresponding author. Tel.: +1 850 644 9363.
E-mail address: kaschak@psy.fsu.edu (M.P. Kaschak).

0022-2860/$ - see front matter
doi:10.1016/j.cognition.2004.06.005

during the comprehender's interaction with the world (Barsalou, 1999). Thus, understanding a sentence such as "Meghan gave me a toy" requires the retrieval of perceptual information to simulate the objects described in the sentence (e.g. the toy; Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002) and the retrieval of motoric information to simulate the "giving" action of the sentence (Glenberg & Kaschak, 2002).

The view that language comprehension involves the recruitment of sensorimotor information has been supported by several recent findings. Neuroimaging studies have revealed that the processing of words recruits the same neural regions that are active when the referents of the words are processed (e.g. Isenberg et al., 1999; Kan, Barsalou, Solomon, Minor, & Thompson-Schill, 2003; Martin & Chao, 2001; Pulvermüller, 1999). Furthermore, behavioral studies have shown that motor information (Glenberg & Kaschak, 2002) and visual information related to the shape and orientation of objects are activated during sentence processing (Stanfield & Zwaan, 2001; Zwaan et al., 2002; see also Boroditsky, 2000; Fincher-Kiefer, 2001; Spivey, Richardson, & Gonzalez-Marquez, in press).

These data converge on the conclusion that the neural machinery responsible for perceiving and understanding events in the world is engaged during language comprehension. However, significant issues remain to be addressed. Chief among them are questions about how exactly the mechanisms involved in perception and motor planning are used in generating a simulation. The most extreme claim is that the exact same mechanisms that are involved in the perception of events (or planning of actions) are also used during the comprehension of language describing those events. If this is the case, simulations generated for sentences describing movements with the hands and fingers (such as playing the piano) should primarily engage motor areas involved in planning hand and finger movements. Similarly, linguistic descriptions of a specific color ("red") should engage the neurons that show maximal responding to that color when it is perceived in distal objects in the environment. The present study takes a first step towards addressing this claim.

Participants engaged in a language processing task while simultaneously viewing a visual display. They listened to sentences describing events that involved movement in one of four directions: towards (e.g. "The car approached you"), away (e.g. "The car left you in the dust"), up (e.g. "The rocket blasted off"), or down (e.g. "The confetti fell on the parade"). They made sensibility judgments (Experiment 1) or grammaticality judgments (Experiment 2) on the sentences. While performing this task, the participants watched a visual display that depicted motion in one of four directions: towards, away, up, or down. The visual stimuli were black-and-white spirals (for towards and away) or black-and-white horizontal bars (for up and down) that created the percept of motion (see Figs. 1 and 2). These stimuli were chosen because they were fairly minimal in terms of content, yet were sufficient to produce the perception of motion in the desired direction. If the comprehension of sentences that describe motion in a particular direction involves the mechanisms responsible for perception of motion, then the comprehension of sentences in these experiments should be affected by concurrently viewing the visual displays.

There are two competing hypotheses regarding the influence of the visual percept on sentence comprehension. The first prediction is that sentences will be processed more easily when the participant is viewing a visual stimulus depicting motion in the direction

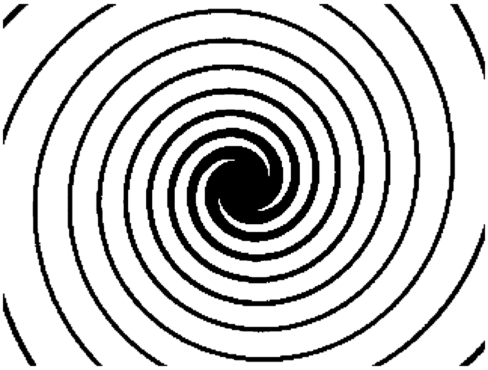

Fig. 1. Spiral stimulus for "towards" and "away" displays.

described in the sentence. For instance, participants should find it easier to understand, "The car approached you," when viewing a display that depicts motion towards them. This prediction was based on the results of previous experiments showing faster responding to visual stimuli that matched the content of a recently processed sentence (Stanfield & Zwaan, 2001; Zwaan, Madden, Yaxley, & Aveyard, 2004; Zwaan et al., 2002). The "Match" advantage arises because the perceptual simulation constructed while comprehending the sentence creates a pattern of activation in visual processing areas that facilitates the processing of visual percepts (such as a picture) matching that pattern of activation (see McDermott & Roediger, 1994, for a similar finding within a memory paradigm).

The second prediction is that sentences will be processed with more *difficulty* when they describe motion in the *same* direction as the motion depicted in the visual display. On this account, perceiving motion in one particular direction engages neurons that respond preferentially to motion in that direction (Mather, Verstraten, & Anstis, 1998). Because these neurons are engaged by the visual stimulus, they are less available for use in constructing a simulation of events in which the action moves in the same direction.

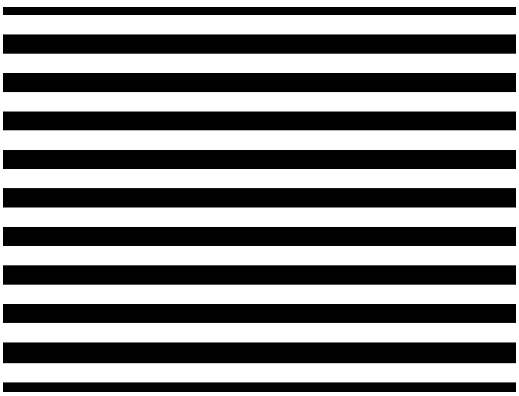

Fig. 2. Horizontal bar stimulus for "up" and "down" displays.

This prediction is supported by experiments that have examined the interplay between visual imagery and visual perception. As one example, participants' ability to make judgments about a mental image of an object is impaired when they are required to simultaneously view a picture of an object of the same semantic category (Lloyd-Jones & Vernon, 2003; see also Craver-Lemley & Reeves, 1987; Perky, 1910). If we consider the perception-action simulation constructed while comprehending language to be akin to visual imagery, it is reasonable to predict that engaging the visual system with an image of motion in one direction should impair one's ability to simultaneously simulate motion in that same direction.

1. Experiment 1

Participants listened to sentences that described motion in a particular direction and had to decide if they were sensible or not. They simultaneously viewed visual displays depicting motion that was either in the same direction as that described in the sentence (Match condition), or in the opposite direction (Mismatch condition). The experiment was designed for two purposes: (1) to determine if sentence processing would be affected by the simultaneous perception of visual motion, and (2) to determine whether sentences would be processed faster when the visual percept matched the direction of the motion in the sentence, or when the visual percept mismatched the direction of the motion in the sentence.

1.1. Method

1.1.1. Participants

Forty-eight introductory psychology students from Florida State University participated in the experiment as part of a course requirement. Five participants were eliminated from the study for having accuracy significantly lower than the other participants in the experiment (accuracy $<80\%$), leaving a total of 43 participants.

1.1.2. Materials

Thirty-two critical sentences were constructed describing motion in one of four directions from the listener: towards, away, upwards, and downwards (see Appendix A for all critical sentences). Eight sentences were written for each direction. In addition, 48 filler sentences were constructed. In order to disguise the critical trials, all the filler sentences described events involving the senses (seeing, hearing, smelling, and so on). None of the filler sentences described events with motion. Twenty of these filler sentences were constructed to be nonsensical (e.g. "The flower sorted the marathon"). All sentences were spoken by a male speaker and recorded on a PC using Cakewalk© Sonar© XL 2.0 with Soundforge© 5.0 plug-ins.

Four black-and-white motion percepts were constructed to accompany the sentences. The "toward" percept was constructed by presenting a black and white spiral picture rotated 9° for every 30 ms presentation. The "away" percept was constructed by presenting the same sequence in the reverse order. The "down" percept was constructed by presenting a block of black-and-white horizontal stripes that moved down by 5 pixels for each 30 ms presentation, and the "up" percept was constructed by presenting the same sequence in the reverse order. Each percept occupied the entire viewable region of the 19 in. flat-screen display at a resolution of 640×480 pixels. All four percepts had a small red fixation cross at the center of the display.

Each percept was presented for 35 s with a 21 s break (white screen) between percepts. A given percept occurred twice during the course of the experiment. During a 35 s percept presentation, 10 sentences were presented over headphones with 3 s between sentence onsets. Two of these sentences matched the concurrently displayed motion percept, two of these sentences mismatched the motion percept, and six were filler sentences. Two or three of the filler sentences during each percept were nonsensical. The 10 sentences for a given percept were presented in a preset random order, identical for each presentation. The order of the eight percepts (four percepts each shown twice) was also a preset randomized order. Four lists were constructed to counterbalance sentence-percept match and sentence-percept order. If a given sentence was presented during a matching percept on list 1 (away sentence — away percept), it was presented during the opposite percept on list 2 (away sentence — towards percept). Lists 3 and 4 were identical to lists 1 and 2 except that the 8 percept presentations were presented in reverse order (the randomized sentence order within each percept was always the same). Each participant was exposed to only one list.

1.1.3. Procedure

Participants were instructed to stare at the fixation cross at the center of each percept presented on the computer screen while listening to sentences over headphones. For each sentence, the participant pressed a key labeled "Y" if the sentence made sense, and a key labeled "N" if the sentence did not make sense. Participants were informed that their responses would be timed, and that they should respond as quickly as possible while still maintaining accuracy. They saw 8 percept presentations, each lasting 35 s. During each percept presentation participants heard and responded to 10 sentences over headphones. After all 10 sentences had been presented, the percept disappeared and a blank screen was displayed for 21 s while the participant took a break. The onset of the next percept cued participants that the next set of sentences was about to start.

1.1.4. Design and analysis

The participants were able to respond to the sentences at any time during the trial. We calibrated the scale of the response times such that a response that occurred at the very end of the sentence would have a value of 0 ms, with positive response times indicating a response after the end of the sentence, and negative response times indicating a response before the end of the sentence. The following procedures were used to screen for outliers. First, based on a visual inspection of the distribution of response times, responses less than -500 ms (i.e. 500 ms before the end of the sentence) and longer than 900 ms were excluded. The visual inspection was aimed at finding large gaps in the distribution of response times, where response times that fell outside of these gaps were deemed to be outliers. Then, for each participant, we removed scores that were more than 2 standard deviations from their mean response time. Finally, all incorrect responses were discarded. This resulted in the loss of 4.43% of the responses. The remaining response times were

Table 1
Mean response times (in ms) and proportion of correct responses for Experiments 1 and 2 (standard error of the mean in parentheses)

	Response time		Accuracy	
	Match	Mismatch	Match	Mismatch
Experiment 1	241 (19)	221 (20)	0.94 (0.01)	0.95 (0.01)
Experiment 2	369 (28)	330 (25)	0.95 (0.01)	0.95 (0.01)

submitted to a 2 (Match vs. Mismatch)×4 (List 1 vs. List 2 vs. List 3 vs. List 4) mixed factor ANOVA with List as a between-subjects variable. We did not include separate percept conditions in the analysis because there were too few items per participant to make meaningful comparisons. We do not report effects involving List because they are of little theoretical interest.

1.2. Results and discussion

The data are presented in Table 1. Analyses denoted $F1$ were conducted with subjects as a random factor in the design, and analyses denoted $F2$ were conducted with items as a random factor. All effects reported as significant are $p<0.05$ unless otherwise noted. Participants responded more quickly to sentences in the Mismatch condition than to sentences in the Match condition [$F1(1, 39)=4.17$, MSe = 2030; $F2(1, 31)=8.19$, MSe = 1239]. There was no evidence of a speed-accuracy trade-off in the experiment, with accuracy being equivalent across experimental conditions [$F1$ and $F2<1$].

The Mismatch advantage supports the claim that there is a high degree of specificity in the extent to which perceptual mechanisms are recruited to conduct simulations during language comprehension. When perceiving a stimulus that engages mechanisms tuned to respond to motion in a particular direction, participants had a more difficult time making sensibility judgments for sentences that described motion in that same direction. However, when perceiving a stimulus that engages mechanisms tuned to respond to motion in a direction opposite to that described in the sentence, participants had a comparatively easy time making the judgment. This result suggests that constructing a simulation of motion during language comprehension requires not only the same general mechanisms that are involved in perception of motion, but also perhaps the specific neurons that are tuned to respond to motion in that direction. When these neurons are engaged by a visual percept, they are less available to aid in the construction of the simulation, leading to slower response times in the Match condition.[1]

[1] As there is no control condition in this experiment, it is unclear whether the Mismatch advantage is due to interference in the Match condition, or priming in the Mismatch condition. Although the available data do not directly address this issue, it seems most plausible to us that the effect is due to interference in the Match condition (along the lines laid out in the introduction). Our preference for the "interference" explanation arises in part because we know of no plausible explanation for why priming would occur in the Mismatch condition.

2. Experiment 2

Experiment 2 was designed to replicate and modestly extend the results of Experiment 1. The methodology was the same as that used in the first experiment, except that participants were asked to make a grammaticality judgment on the test sentences. The purpose of this manipulation was to examine if the Mismatch advantage could be observed in task conditions that do not emphasize semantic processing. If the effect observed in Experiment 1 is replicated, it would lend more credence to the claim that perceptual simulation routinely occurs during sentence processing.

2.1. Method

2.1.1. Participants

Thirty-seven introductory psychology students from Florida State University participated in the experiment as part of a course requirement. One participant was eliminated from the study for having accuracy significantly lower than the other participants in the experiment (accuracy $<75\%$), leaving a total of 36 participants.[2]

2.1.2. Materials

The critical sentences were the same as in Experiment 1. The filler sentences were also the same, except that the 20 nonsensical sentences were replaced with 20 ungrammatical sentences.

2.1.3. Procedure

The procedure was the same as for Experiment 1, except that participants made grammaticality judgments on the sentences.

2.1.4. Design and Analysis

The response times were collected and screened for outliers as in Experiment 1. The only difference in the methods was that our visual inspection of the data led us to select a slightly higher cutoff for long response times than in the first experiment (1750 ms). These procedures resulted in the loss of 1.9% of the observations. The analyses conducted on the remaining response times were identical to those used in Experiment 1.

2.2. Results and discussion

The results are presented in Table 1. Similar to Experiment 1, analyses revealed a Mismatch advantage [$F1(1, 32)=3.90$, $MSe=6892$, $p=0.057$; $F2(1, 31)=5.79$, $MSe=5557$]. Accuracy did not differ across conditions [$F1$ and $F2<1$]. Thus, these data

[2] Our cutoff for accuracy changed between Experiment 1 and Experiment 2 because the distribution of accuracy scores in each experiment was slightly different (ostensibly due to differences in the tasks). This lead to a slightly lower cutoff in Experiment 2.

demonstrate that the effect can be observed in a language comprehension task that does not emphasize semantic processing.

3. General discussion

The Mismatch advantage observed in these experiments presumably reflects a conflict between the demands of processing the visual percept and the demands of constructing a simulation during language comprehension. Perception of motion in the visual display engages neural mechanisms tuned to respond to motion in a particular direction. Constructing a simulation of events that incorporate motion in that same direction also requires the use of the mechanisms tuned to respond when motion is perceived. Because these neural mechanisms are already engaged by the visual percept, they are less available for use in constructing the simulation, hindering participants' ability to comprehend the sentence. When the direction of the visual percept does not match the direction of motion in the sentence, no such conflict arises, and sentence comprehension can proceed more easily. This demonstration of selective interference between a visual processing task and a language comprehension task both provides further support for the claim that language comprehension is grounded in perception and action, and clarifies in part how this grounding occurs. The data suggest that there is considerable specificity in the simulations constructed during language comprehension. Perceiving motion did not simply interfere with the comprehension of language about motion; rather, perceiving motion in a particular direction selectively impaired processing of sentences that describe motion in that same direction.

It is worth noting the apparent disconnect between earlier studies showing that responding is faster when perceptual stimuli or motor responses match the content of the sentence (e.g. Glenberg & Kaschak, 2002; Zwaan et al., 2002), and the present Mismatch advantage. Two factors interact in determining whether a Match or Mismatch advantage should be found in tasks such as these: temporal overlap and integratability (see Kosslyn, 1994; Lloyd-Jones & Vernon, 2003, for similar proposals regarding the presence of interference or facilitation between mental imagery and visual perception). Temporal overlap is the relative timing of the perceptual stimulus and the sentence to be processed. Integratibility refers to the extent to which the perceptual stimulus can be integrated into the simulation constructed of the content of the sentence. For example, a picture of a car can be integrated with the content of the sentence, "The car approached you," whereas the black-and-white displays used in these experiments cannot.

When the sentence and the perceptual stimulus must be processed at the same time (as in the experiments reported here), a Match or Mismatch advantage may occur, depending on the extent to which the sentence and the stimulus are integratible. If the sentence and the stimulus are integratible (e.g. when one sees the image of a car while processing the sentence, "The car approached you"), a Match advantage is expected. Whereas the image of a car approaching you (Match) can easily be integrated into (or, prime comprehension of) the content of a sentence about an approaching car, the image of a car moving away from you (Mismatch) cannot, resulting in comprehension difficulty. Contrariwise, if the sentence and the stimulus are generally not integratible (e.g. the black-and-white stimuli used in our experiments are not integratible with the content of the critical sentences we used), a Mismatch advantage is expected. This is presumably because the perceptual stimulus engages the processing mechanisms needed to simulate the sentence, which results in comprehension difficulty when the content of the sentence matches the content of the stimulus (as discussed above).

When the perceptual stimulus and sentence are processed sequentially (i.e. when there is no temporal overlap), there should be a Match advantage when the percept and the content of the sentence are integratible. The Match advantage is the result of priming or "tuning" within the perceptual system, as reported in various non-linguistic tasks (e.g. Farah, 1985; McDermott & Roediger, 1994). As before, the image of an approaching car should prime the comprehension of a sentence about an approaching car, whereas the image of a car moving off into the distance should slow comprehension of a sentence about an approaching car. When the percept and the content of the sentence are not integratible, there should be a null effect (i.e. the stimulus should have little or no effect on the processing of the sentence). The null effect is observed because the perceptual mechanisms can process the percept and the simulation of the sentence independently due to the lack of temporal overlap and shared content between them.

Further exploration of this hypothesis requires the development of a precise metric of integratibility. Whereas it may be intuitive that the image of a car should be integratible with the simulation of a sentence about a car, and that the black-and-white percepts used in these experiments should not, there are many cases where the integratibility or non-integratibility of stimuli is harder to determine. For example, to what extent is a picture of a Model-T Ford integratible with a sentence about a BMW driving down the street? Furthermore, is integratibility an all-or-none proposition? Findings from the literature on the interaction between imagery and perception may provide some clues in this regard (see Lloyd-Jones & Vernon, 2003), but it remains to be seen how the variables implicated in this work will play out in the domain of sentence comprehension. Definitive answers to these questions must await further empirical investigation.

Acknowledgements

Michael P. Kaschak, Carol J. Madden, David J. Therriault, Richard H. Yaxley, Mark E. Aveyard, and Rolf A. Zwaan (Department of Psychology) and Adrienne Blanchard (Department of Communicative Disorders), Florida State University.

This research was supported by grant MH-63972 to Rolf Zwaan. The authors would like to thank Nicole Botero, Cassandra Douglas, Lisa Matlock, and Michael Anderson for running the experiments reported here. Thanks also to Michael F. Neelon for helpful discussions of the perception of visual motion.

Appendix A. Critical sentences from Experiments 1 and 2

The critical sentences used in Experiments 1 and 2 are presented below. They are sorted by their direction of motion. Filler sentences are available upon request from the first author.

Away:
He rolled the bowling ball down the alley.
You backed out of the driveway.
The horse ran away from you.
You backed away from the fire.
The car left you in the dust.
The squirrel scurried away.
You pushed off from the dock.
The train pulled away from you.
Towards:
The dog was running towards you.
You ran towards the door.
The storm clouds rolled in.
You gained upon the runner.
You closed in on the ball.
The car approached you.
The shark was drawing near you.
The buffalo charged at you.
Up:
The steam rose from the boat.
The smoke rose into the sky.
The cat climbed the tree.
The balloon ascended into the clouds.
The referee tossed up the basketball.
You raised the flag.
The curtain went up.
The fireworks shot up.
Down:
The sand poured through the hourglass.
The eagle dove to catch its prey.
The leaves fell from the tree.
You lowered the blinds.
The snow fell onto the ground.
The confetti rained down on the parade.
The water dripped from the faucet.

References

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

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75*, 1–28.

Craver-Lemley, C., & Reeves, A. (1987). Visual imagery selectively reduces Vernier acuity. *Perception*, *16*, 533–614.

Farah, M. J. (1985). Psychophysical evidence for a shared representational medium for visual images and percepts. *Journal of Experimental Psychology: General*, *114*, 93–105.

Fincher-Kiefer, R. (2001). Perceptual components of situation models. *Memory and Cognition*, *29*, 336–343.

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

Isenberg, N., Silbersweig, D., Engelien, A., Emmerich, K., Malavade, K., Beati, B., Leon, A. C., & Stern, E. (1999). Linguistic threat activates the human amygdala. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 10456–10459.

Kan, I. P., Barsalou, L. W., Solomon, K. O., Minor, J. K., & Thompson-Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology*, *20*, 525–540.

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

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Lloyd-Jones, T. J., & Vernon, D. (2003). Semantic interference from visual object recognition on visual imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 563–580.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology*, *11*, 194–201.

Mather, G., Verstraten, F., & Anstis, S. (Eds.). (1998). *The motion aftereffect: A modern perspective*. Cambridge, MA: MIT Press.

McDermott, K., & Roediger, H. (1994). Effects of imagery on perceptual implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 418–434.

Perky, C. W. (1910). An experimental study of imagination. *Journal of Psychology*, *21*, 422–452.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253–279.

Spivey, M. J., Richardson, D. C., & Gonzalez-Marquez, M. (in press). On the perceptual-motor and image-schematic infrastructure of language. In D. Pecher & R. A. Zwaan (Eds.), *The grounding of cognition: The role of perception and action in memory, language, and thinking*. Cambridge, UK: Cambridge University Press.

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

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H Ross, *Psychology of Learning and Motivation* (Vol. 44) (pp. 35–62). San Diego, CA: Academic Press.

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.

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