Language comprehenders represent object distance both visually and auditorily

BODO WINTER[#] AND BENJAMIN BERGEN[+*]

[#]*University of California, Merced*
[+]*University of California, San Diego*

Abstract

When they process sentences, language comprehenders activate perceptual and motor representations of described scenes. On the "immersed experiencer" account, comprehenders engage motor and perceptual systems to create experiences that someone participating in the described scene would have. We tested two predictions of this view. First, the distance of mentioned objects from the protagonist of a described scene should produce perceptual correlates in mental simulations. And second, mental simulation of perceptual features should be multimodal, like actual perception of such features. In Experiment 1, we found that language about objects at different distances modulated the size of visually simulated objects. In Experiment 2, we found a similar effect for volume in the auditory modality. These experiments lend support to the view that language-driven mental simulation encodes experiencer-specific spatial details. The fact that we obtained similar simulation effects for two different modalities—audition and vision—confirms the multimodal nature of mental simulations during language understanding.

Keywords
embodied cognition, mental simulation, language comprehension, distance, immersed experiencer

* Correspondence address: Bodo Winter, Department of Cognitive and Information Sciences, University of California at Merced, 5200 North Lake Road, Merced, CA 95343, USA. E-mail: bodo@bodowinter.com. We thank Teenie Matlock, Victoria Anderson, Joelle Kirtley, Clive Winter, Amy Schafer, members of the PIG group, and participants of LING 750X for helpful comments and suggestions.

Language and Cognition 4–1 (2012), 1–16
DOI 10.1515/langcog-2012-0001

1866–9808/12/0004–0001
© Walter de Gruyter

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

1. Introduction

Converging evidence from behavioral experimentation and brain imaging suggests that language comprehenders construct mental simulations of the content of utterances (for reviews, see Bergen 2007; Barsalou et al. 2008; Taylor and Zwaan 2009). These mental simulations encode fine perceptual detail of mentioned objects, including such characteristics as motion (Kaschak et al. 2005), shape (Zwaan et al. 2002), orientation (Stanfield and Zwaan 2001), and location (Bergen et al. 2007). Some researchers have taken these findings to suggest that comprehenders construct mental simulations in which they virtually place themselves inside described scenes as "immersed experiencers" (Barsalou 2002; Zwaan 2004). This "immersed experiencer" view argues that understanding language about a described scene is akin to perceptually and motorically experiencing that same scene as a participant in it. As a result, objects mentioned in sentences ought to be, on this view, mentally simulated as having perceptual properties reflecting the viewpoint that someone immersed in the scene would take—reflecting, for instance, angle and distance.

However, it is equally plausible that language processing engages perceptual and motor systems without rendering described scenes from a particular, immersed perspective. The human vision system encodes viewpoint-invariant representations of objects (Vuilleumier et al. 2002) that could in principle be recruited for understanding language about objects. In fact, nearly all of the current evidence that language comprehension engages motor and perceptual systems (with a few exceptions discussed below) is consistent with both the "immersed experiencer" and this alternative, "viewpoint-invariant" possibility. For instance, experimental results showing that people are faster to name an object when it matches the shape implied by a preceding sentence (an egg "in a pot" or "in a skillet," for instance) do not reveal whether the comprehender represents the object as seen from a particular viewpoint or distance (Zwaan et al. 2002).

A number of existing studies support the "immersed experiencer" view of mental simulation. For one, Yaxley and Zwaan (2007) demonstrated that language comprehenders mentally simulate the visibility conditions of described scenes: after reading a sentence such as *Through the fogged goggles, the skier could hardly identify the moose*, participants responded more quickly to a blurred image than to a high resolution image of a mentioned entity (such as a moose). Horton and Rapp (2003) and Borghi et al. (2004) provided similar findings for the simulation of visibility and accessibility. Finally, several studies have found that personal pronouns such as *you* or *she* can modulate the perspective of a mental simulation (Brunyé et al. 2009; Ditman et al. 2010).

The present work tests two different predictions of the immersed experiencer view. First, if comprehenders simulate themselves as participants in de-

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

scribed scenes, then linguistic information about the distance of objects from a perceiver should modulate the perceptually represented distance of simulated objects. Early work on situation models has shown that distance information can affect language understanding (for a review, see Zwaan and Radvansky 1998). For example, comprehenders exhibit slower recognition times with words that denote objects far away from the protagonist of a story and faster recognition times with close objects (Morrow et al. 1987). However, one limitation of many previous studies on distance is that participants often have extensive training on the spatial set-up of a described situation prior to the language task, for instance, participants often have to memorize items on a map before the actual language task. This means that any effects of distance are produced not by language but through prior visual experience. This leaves open the question whether explicit or implicit distance encoded by language results in perceptually different mental representations of objects. In addition, previous studies using a situation models perspective are ambiguous as to the representational format of the different representations for nearer or farther objects. For example, faster responses to nearby than far-away objects could simply derive from different degrees to which nearby and far-away objects are held active in short-term memory. The experiments described below were designed to both test for effects of linguistic manipulations of distance and do so in a way that directly assesses the perceptual characteristics of the resulting representations.

The second prediction we test is the claim that mental simulations are multimodal, including not only visual but also auditory characteristics of mentioned objects (Barsalou 2008; Barsalou 2009; Taylor and Zwaan 2009). While there has been a good deal of work on visual simulation (Stanfield and Zwaan 2001; Zwaan et al. 2002; Kaschak et al. 2005) and motor simulation (Glenberg and Kaschak 2002; Bergen and Wheeler 2005; Bergen and Wheeler 2010; Wheeler and Bergen 2010) in language understanding, very little work has addressed language-induced auditory simulation (Kaschak et al. 2006; van Danzig et al. 2008; Vermeulen et al. 2008). The current experiment complements this experimental work by demonstrating a compatibility effect that is created by the simulation of distance in both the visual and the auditory modality. Since language-induced mental simulation is frequently claimed to be multimodal, it seems desirable to find more evidence for the presence of language-induced auditory simulation, and to find evidence for not only the auditory simulation of motion (Kaschak et al. 2006), but also for the simulation of other spatial features such as distance. Distance has previously only been considered with respect to map-based tasks in which the auditory modality did not play a role.

In two experiments, we examine effects of the distance of mentioned objects on comprehender simulation in two modalities. A sentence like *You are looking*

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

at the milk bottle across the supermarket (Far condition) should lead one to simulate a smaller milk bottle than the sentence *You are looking at the milk bottle in the fridge* (Near condition). Likewise, a language comprehender should simulate a quieter gunshot upon reading *Someone fires a handgun in the distance* (Far condition), and a louder one when reading *Right next to you, someone fires a handgun* (Near condition). Crucially, the two experiments are very similar with respect to their designs and thus allow us to test predictions of the immersed experiencer account in the visual and the auditory modalities using similar metrics.

2. Experiment 1: Visual distance

2.1. *Design and predictions*

The design we employed is a variant of the sentence-picture matching task first used by Stanfield and Zwaan (2001) and later adopted by Zwaan et al. (2002), Yaxley and Zwaan (2007) and Brunyé et al. (2009), among others. Participants read sentences and subsequently saw pictures of objects that were either mentioned in the sentence or not. The participant's task was to decide whether the object was or was not mentioned in the sentence. In all critical trials, the object had been mentioned in the preceding sentence. The reasoning underlying this task is that reading a sentence should lead the reader to automatically perform a mental simulation of its content. The more similar a subsequent picture is to the reader's mental simulation, the more should responses be facilitated (Bergen 2007). In a 2 × 2 design, we manipulated the object distance implied by the sentence (Near vs. Far) and the size of the picture (Large vs. Small). The immersed experiencer account predicts an interaction between Sentence Distance and Picture Size—response latencies should be faster when the distances implied by the sentence and the picture match.

2.2. *Materials*

We constructed 32 critical sentence-picture pairs, all of which required yes-responses. To induce the perspective of a participant in the simulated scenes rather than the perspective of an external observer, the subject of all sentences was the pronoun *you* (Brunyé et al. 2009). In addition, all sentences were presented with progressive grammatical aspect, because previous work has shown that progressive grammatical aspect leads participants to adopt an event-internal perspective of simulation (Bergen and Wheeler 2010). All verbs were verbs of visual perception (e.g. *looking*, *seeing*).

Half of the critical sentences marked distance through prepositional phrases or adverbials like *a long way away* or *close to you*, which identified the object's location by implicitly or explicitly referring to the protagonist's location

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

(protagonist-based stimuli). The other half employed prepositional phrases or adverbials that located the object with respect to other landmarks (landmark-based stimuli), e.g. *a frisbee in your hand* versus *a frisbee in the sky*. In addition, we included 32 form-similar filler sentence-picture pairs that required no-responses. To ensure that participants would pay attention to the landmark, we included 32 additional fillers, where the picture following the sentence either matched or mismatched the landmark in the prepositional phrase. To distract participants from the purpose of the experiment, we also included 112 fillers about left/right orientation such as *You are placing the ace of spades to the left of the queen of hearts*. Half required yes-responses, and half no-responses.

We created visual representations of objects to go with each sentence. The objects were all "token invariant" (Haber and Levin 2001)—objects that in the real world display relatively little variation in size across exemplars. We did this to avoid the possibility that the near and far pictures could be mistaken for large and small tokens of the same object. To create the two images for each object, we took a single image and manipulated its size (Small: 200 px vs. Large: 800 px on the longest axis, 72 dpi), sharpness (Gaussian blur filter with 0.3 px radius on Small pictures), contrast (Small: −12 on the contrast scale of Adobe Photoshop) and illumination (Small: +4 on the illumination scale) to create Large and Small versions of the pictures. The images in the 112 left-to-right fillers varied in size between 100 px and 900 px in order to distract from the fact that the critical items appeared in only two sizes.

2.3. *Procedure*

The procedure was managed by E-Prime Version 1.2 (Schneider et al. 2002). In each trial, participants read a sentence, then pressed the spacebar to indicate they had read it. Then a fixation cross appeared for 250 ms, followed by a picture. Participants indicated whether the picture matched the preceding sentence by pressing "Yes" (the "j" key on the keyboard) or "No" ("k"). There were eight practice trials during which the experimenter was present to answer questions. The practice trials included accuracy feedback—in the actual experiment, there was no feedback. After half of the stimuli were presented in the actual experiment, participants were given an optional break.

2.4. *Participants*

Twenty-two undergraduate students of the University of Hawai'i at Mānoa received credit for an undergraduate linguistics course or small gifts for their participation. All were native speakers of English and reported normal or corrected-to-normal vision.

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

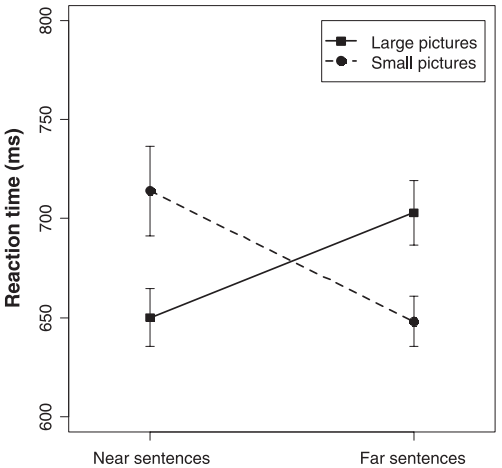


Figure 1. *Reaction times to Large pictures and Small pictures depending on Sentence type (Near vs. Far); bars represent standard errors.*

2.5. *Results*

All participants performed with high mean accuracy ($M = 97\%$, $SD = 0.04\%$); none were excluded on the basis of accuracy. We excluded inaccurate responses (2.8% of the data) and winsorized remaining response times over 3 standard deviations from each participant's mean (we replaced values exceeding 3 standard deviations with the maximum value of each participant that is within 3 standard deviations, this affected 2.5% of the remaining data; see Barnett and Lewis 1978).

We performed two two-way repeated-measures ANOVAs with Sentence Distance (Near vs. Far) and Picture Size (Large vs. Small) as fixed factors, and participants (F_1) and items (F_2) as random factors. There were no main effects ($Fs < 1$), however, there was a significant interaction of Sentence Distance and Picture Size by participants ($F_1(1,21) = 6.14$, $p = 0.02$, $\eta_p^2 = 0.230$) and items ($F_2(1,31) = 4.54$, $p = 0.04$, $\eta_p^2 = 0.126$). Response latencies were on average 649 ms in the matching and 709 ms in the mismatching condition (a 60 ms difference; see Figure 1). A separate $2 \times 2 \times 2$ ANOVA with distance cue (Protagonist-based vs. Landmark-based) as an additional fixed factor tested whether having protagonist-based or landmark-based linguistic distance cues affected the results. There was no significant three-way interaction by subjects or by items ($Fs < 1$), thus indicating that it did not.

Post-hoc pairwise comparisons showed that there was a difference between Large and Small pictures for Far sentences by subjects ($t_1(21) = 2.448$, $p = 0.023$; $t_2(31) = 1.708$, $p = 0.098$), as well as between Far and Near sentences

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

for Large pictures by subjects ($t_1(21) = 2.372$, $p = 0.027$; $t_2(31) = 1.322$, $p = 0.196$), however, these are not significant by the Bonferroni-corrected alpha level of 0.008.

In order to test for a possible speed-accuracy trade-off, we conducted ANOVAs on mean accuracy per condition. There was no indication of a speed-accuracy trade-off; participants were somewhat more likely to respond correctly in the (faster) matching conditions (98% vs. 96% mean accuracy). This trend was significant by items ($F_2(1,31) = 4.613$, $p = 0.04$, $\eta_p^2 = 0.130$) but not by participants ($F_1(1,21) = 1.846$, $p = 0.19$, $\eta_p^2 = 0.081$).

These results support the first prediction made by the immersed experiencer account. When reading sentences about distant objects, comprehenders simulate smaller objects, and when they read sentences about nearby objects, they simulate larger objects. This effect occurred regardless of whether distance was protagonist-based or landmark-based.

Experiment 2 explores parallel effects of distance in auditory simulation, testing a second prediction of the immersed experiencer account—that language-driven mental simulation is multi-modal. Experiment 2 also addresses a possible concern with the sentence materials that were presented in Experiment 1. All the verbs were verbs of visual perception. This might have artificially caused participants to focus on distance, perhaps because sentences like *You are looking at the living room door* could be interpreted as instructions for conscious mental imagery. To deal with this issue, the sentence materials of Experiment 2 are less explicit and do not use verbs of perception.

3. Experiment 2: Auditory distance

3.1. *Design and predictions*

Where Experiment 1 employed a sentence-picture matching task, Experiment 2 implemented a sentence-sound matching task. Participants read sentences and subsequently heard sounds of objects or animals that were either mentioned in the sentence or not. The participant's task was to verify whether the sound they heard was of an entity mentioned in the sentence. We manipulated Sentence Distance (Near, Far) and Sound Volume (Loud, Quiet) in a 2 × 2 design. The multimodal component of the immersed experiencer view predicts an interaction between the two factors.

3.2. *Materials*

Twenty-four critical sentence pairs described an entity as Near to or Far from the event participant; all required yes-responses. We also included 24 no-response sentences as fillers. Sentences were in the present tense or present progressive. For each pair of critical sentences, we constructed corresponding

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

Loud and Quiet sounds (quantization: 16 bit; sampling frequency: 22,050 Hz). We began with a single sound for each pair of sentences. We then manipulated amplitude and spectral slope with Audacity and Praat (Boersma and Weenink 2009). Increasing the distance of a sound source by 10 meters leads to a decrease of approximately 20 dB in intensity (Zahorik 2002: 1837), so we manipulated stimuli such that they had an average intensity of 60 dB in the Loud condition and 40 dB in the Quiet condition. In addition, when sounds are propagated over long distances, higher frequencies are dampened more than lower frequencies (Ingard 1953; Coleman 1968). We thus applied a filter to the Quiet sounds that reduced frequencies above 1 kHz by 4.5 dB per octave. To confirm that the difference between Loud and Quiet sounds was audible, we performed a norming study with 10 participants who were asked to indicate which of two versions of a sound played in sequence was perceived as being "closer". Participants were on average 97% correct in deciding whether a sound was near or far, indicating that the distance manipulation is in fact audible and easy to perceive.

3.3. *Procedure*

The procedure was similar to experiment 1. Visual sentence presentation ended when the participant pressed the space bar. A blank screen then appeared for 200 ms, followed by a sound. On 25% of the trials, comprehension questions followed the sounds, in order to ensure that participants attended to the entire sentence. Again, there were eight practice trials with feedback before the main experiment.

3.4. *Participants*

Thirty-three undergraduates at the University of Hawai'i at Mānoa participated in the experiment and received course credit or small gifts for participating. All were native speakers of English, who reported normal or corrected-to-normal vision and hearing.

3.5. *Results*

One participant was excluded because his accuracy on sound verification was below 80% (all other participants: $M = 95\%$, $SD = 0.05\%$). Inaccurate responses (4.9%) were excluded, and trials that deviated by more than 3SDs from each participant's mean were windsorized (1.9%). Repeated-measures ANOVAs revealed main effects of Sound Volume and Sentence Distance. Loud sounds lead to faster verification times by both participants and items ($F_1(1,31) = 10.7$, $p = 0.003$, $\eta_p^2 = 0.257$; $F_2(1,23) = 11.69$, $p = 0.002$, $\eta_p^2 = 0.337$); Near sentences lead to faster verification times by participants

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

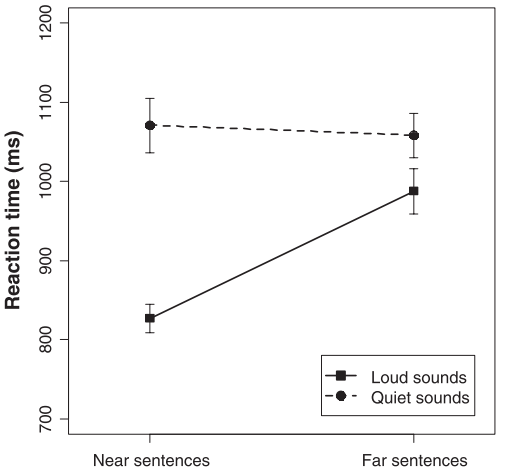


Figure 2. *Reaction times to Loud sounds and Quiet sounds depending on Sentence type (Near vs. Far); bars represent standard errors.*

($F_1(1,31) = 5.4$, $p = 0.03$, $\eta_p^2 = 0.149$) but not by items ($F_2(1,23) = 2.89$, $p = 0.1$, $\eta_p^2 = 0.112$). Crucially, we obtained the predicted interaction between Sentence Distance and Sound Volume by participants ($F_1(1,31) = 5.3$, $p = 0.03$, $\eta_p^2 = 0.147$) and a marginally significant interaction by items ($F_2(1,23) = 4.04$, $p = 0.056$, $\eta_p^2 = 0.149$). In the matching condition, participants' response latencies were on average 942 ms, whereas in the mismatching condition they were 1029 ms (a 87 ms difference; see Figure 2).

Post-hoc pairwise comparisons (Bonferroni-corrected) showed that there was a significant difference between Near sentences and Far sentences for Loud sounds by subjects ($t_1(31) = 3.578$, $p < 0.001$; $t_2(23) = 1.355$, $p = 0.189$), and between Loud sounds and Quiet sounds for Near sentences ($t_1(31) = 3.729$, $p < 0.001$; $t_2(23) = 1.566$, $p = 0.131$). In addition, the difference between the Far/Quiet and the Near/Loud condition was significant by subjects ($t_1(31) = 5.209$, $p < 0.0001$; $t_2(23) = 0.007$, $p = 0.994$).

An error analysis revealed a main effect of sound ($F_1(1,34) = 4.904$, $p = 0.034$, $\eta_p^2 = 0.126$; $F_2(1,23) = 4.570$, $p = 0.043$, $\eta_p^2 = 0.166$); participants were slightly more likely to respond accurately to Loud sounds than to Quiet sounds (96% vs. 93%). Since Quiet sounds also lead to slower response times, this pattern goes into the opposite direction of a speed-accuracy trade-off (similar to the accuracy results in Experiment 1). Crucially, the accuracy data did not reveal an interaction between Sentence Distance and Sound Volume ($F_1(1,34) = 2.280$, $p = 0.140$, $\eta_p^2 = 0.063$; $F_2(1,23) = 1.5$, $p = 0.233$, $\eta_p^2 = 0.061$) and thus there was no indication of a speed-accuracy trade-off.

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

In sum, we found two main effects and an interaction. With respect to the main effects, it is not surprising that Loud sounds were processed faster than Quiet sounds overall, since more intense auditory stimuli generally lead to faster neural response latencies (Sugg and Polich 1995). The finding that sentences describing near objects result in faster responses is, to our knowledge, novel. However, it seems noteworthy in this respect that Sereno et al. (2009) found that words denoting large objects are processed faster than words denoting small objects. Sentences describing near sound sources might be processed faster because they are ecologically more important (for instance, near objects are more relevant for action than far-away objects), or because loud sounds are simulated more easily than quiet sounds, just like loud sounds are perceived faster than quiet sounds (Sugg and Polich 1995).

The interaction effect we observed shows that linguistic information about distance modifies the details of auditory mental simulations, a prediction made by a multimodal version of the immersed experiencer hypothesis. However, in contrast to the results of Experiment 1, this interaction was carried predominantly by the Loud sounds. Given that participants had significantly lower accuracies when responding to Quiet sounds, as well as overall slower response times (157 ms slower than responses to Loud sounds), we suspect that participants generally experienced greater difficulty in responding to the Quiet sounds. This may have masked the effect of Sentence Distance for Quiet sounds.

4. General discussion

Processing sentences about entities close to an event participant leads to faster responses to large, loud representations of those entities, as contrasted with entities far from an event participant, which facilitate responses to small, quiet representations. These results have three implications. (1) Comprehenders perceptually represent distance of mentioned objects. Like other work showing that motion (Kaschak et al. 2005), object orientation (Stanfield and Zwaan 2001), object shape (Zwaan et al. 2002), visibility conditions (Yaxley and Zwaan 2007; Horton and Rapp 2003) and perspective (Brunyé et al. 2009) are relevant dimensions of visual mental simulation, this finding confirms a potentially falsifiable prediction of the immersed experiencer view of mental simulation. If comprehenders simulate the experience of "being there" in mental simulations (Barsalou 2002), they experience specific distances to the objects present in the described scenes. However, it should be pointed out that amodal accounts of language comprehension (see e.g. Mahon and Caramazza 2008) can in principle accommodate our findings post hoc (see discussion in Glenberg and Robertson 2000). From this perspective, our findings are simply the result of spreading activation from language brain areas to the sensory-motor

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

system, the result of a downstream part of the comprehension process that might not play a functional role in language understanding. Our findings do not allow us to conclude that perceptual representations of distance are *necessary* for understanding language about distance—this could only be shown through other methods. (2) Immersing oneself in described experiences entails not only visual simulation, but simulation across relevant modalities. Experiment 2 demonstrates that distance leads to effects on auditory simulation in line with those in vision. This is an important finding because of the scarcity of studies dealing with the auditory modality in language-induced mental simulation. (3) With respect to distance, our results demonstrate that mental simulation is structured similarly to actual perception (Kosslyn et al. 2001). Woodworth and Schlosberg (1954: 481) note that in actual perception, we do not perceive "free-floating objects at unspecified distances," and results from the two experiments above suggest that the same applies to mental simulation.

One important concern with work of this type is that the results might be due to task demands. Perhaps participants were encouraged to perform detailed mental simulations because they saw pictures or heard sounds in each trial. If correct, this criticism affects the external validity of the results reported above. However, there are several reasons to think that the results are not simply due to task demands. First, a number of studies have discovered that effects initially found in sentence-picture matching tasks like the ones conducted here are also present in paradigms that remove images from the experimental design. For example, Ditman et al. (2010) and Pecher et al. (2009) have found that perspective, object shape, and object orientation implied by sentences lead to differences in memory tasks which did not use pictures during the sentence presentation component. Second, a response strategy in which the participant actively generates mental images was discouraged because half of the time, the sentence-picture or the sentence-sound pairs mismatched (for a similar argument, see Stanfield and Zwaan 2001). An active imagery strategy would not improve performance on the task; it might actually hinder it. Finally, in Experiment 1, nearly all (92%) of the sentences mentioned two objects (in the case of the filler items, there were always two objects; in the case of the landmark-based sentences, there always was an object and a landmark), but the pictures only depicted one object. Since the first or the second noun was equally likely to occur in the following picture, participants could not have predicted which object would be depicted in the picture. In addition, the variety of picture sizes in the filler items should have discouraged any distance-related simulations solely due to task-based strategies. For these reasons, it seems unlikely that the results are only due to task demands or top-down response strategies. These experiments are more likely tapping into unconscious and automatic simulation rather than into conscious and purposeful generation of mental imagery.

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

Another possible concern is the use of *you* in the stimuli of both experiments. One might argue that using *you* could have made "immersing oneself in the situation" more likely than it would be with third (or first) person pronouns. But this isn't consistent with previous work, which shows that personal pronouns modulate the perspective on an event that comprehenders adopt. Using *you* is more likely to induce a participant perspective, while third person pronouns are more likely to invite a third person perspective (Brunyé et al. 2009). Critically, comprehenders appear to adopt an immersed perspective with or without the use of *you*; what *you* does is to increase the likelihood of a participant perspective. Because we wanted to manipulate the distance an object would be from the perspective of a particular participant described in the sentence, it was critical that we make consistent use of a single person, and second person allowed for a more systematic manipulation of distance than third person would. However, we hope that future work will investigate the effects of different pronouns on distance effects.

To conclude, we have shown that linguistic information about distance alters the content of mental simulation, which lends support to the view that when constructing mental simulations during language comprehension, we immerse ourselves in detailed situation models that encode perspective-specific spatial relations. Crucially, we have shown that this detailed simulation does not only encompass the visual modality, but also the auditory modality, which supports the idea that mental simulation is multimodal, just like actual perception. This makes understanding language about a scene quite a lot like being in that particular scene.

Appendix: Experimental stimuli

Experiment 1: Landmark-based

You are staring at the living room door from the sofa / across the hallway.
You are staring at the file cabinet in the office / on the far shelf.
You are looking at the baseball bat in your duffle bag / lying on the other side of the field.
You are looking at the milk bottle in the fridge / across the supermarket.
You are eyeing the axe in the tool shed / strewn at the far end of the forest floor.
You are looking at the beer bottle in your fridge / on the end of the counter.
You are eyeing the guitar in the recording room / on the other side of the stage.
You are looking at the violin on this side of the stage / on the other side of the stage.
You are looking at the iPod in your hand / on the other side of the Apple store.

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

You are staring at the frisbee in your hand / in the sky.
You are looking at the shampoo bottle in your bathroom / on the far shelf.
You are looking at the briefcase in your bedroom / at the end of the hallway.
You are looking at the olive oil in the kitchen cabinet / on the far shelf.
You are staring at the exit sign in the classroom / at the far end of the shopping mall.
You are looking at the microphone in your hand / on the other side of the stage.
You are staring at the golf ball in your hand / in the sky.

Experiment 1: Protagonist-based

You are looking at the fire hydrant in front of you / from afar.
You are looking at your bike parked nearby / far away.
You are looking at the bowling ball in front of of you / from afar.
You are staring at the chair next to you / from a distance.
You are looking at the police car parked next to you / far away from you.
Your eyes are fixed on the F1 racing car parked nearby / in the distance.
You are staring at the stop sign in front of you / from a distance.
You are staring at the no-smoking sign in front of you / from a long way away.
You are looking at the lawn-mower nearby / from a distance.
You are staring at the basketball lying next to you / a long way off.
You are looking at the water dispenser in front of you / from a distance.
Your eyes are fixed on the Harley Davidson parked nearby / far away from you.
You are looking at the Coke can in front of you / from a distance.
You are staring the sunflower in front of you / in the distance.
You are staring at the park bench in front of you / at the far park bench.
You are looking at the wheelchair in front of you / which is far away from you.

Experiment 2: Critical stimuli

Near Sentences	Far Sentences
Right next to you, someone fires a handgun.	Someone fires a handgun in the distance.
In the crib right in front of you, there's a baby crying.	In the day-care center down the hall, there's a baby crying.
In the kitchen, you're using the blender to make a smoothie.	You're woken up by your mom downstairs using the blender.
As you are petting the cat, it meows.	A cat somewhere in your neighbor's yard meows.
You hold the champagne bottle in your hand and pop it open.	At the opposite end of the restaurant, someone pops a champagne bottle up.

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

While you're touring the bell tower, the church bells start to ring.	In the neighboring town, the church bells start to ring.
While you're milking the cow, it starts mooing.	Across the field, the cow starts mooing.
You are standing right in the middle of the applauding audience.	From outside, you know the concert is over because the audience is applauding.
The cuckoo-clock right above you strikes midnight.	The cuckoo-clock up in the attic strikes midnight.
You step into the chicken coop and a rooster crows.	Early in the morning, the rooster down the hill crows.
Right next to you, the dog is barking.	In your neighbor's yard, a dog is barking.
You are drilling a screw into the wall with the power drill.	The construction worker across the street is using a power drill.
You are using a hammer to pound a nail into the wall.	A construction worker down the hall pounds a nail into the wall.
The Harley Davidson right in front of you is rumbling.	Blocks away, a Harley Davidson is rumbling.
While you're horseback-riding, your horse neighs.	At the other end of the field, a horse neighs.
You are standing next to a construction worker using a jackhammer.	Somewhere far away from you, a construction worker is using a jackhammer.
As you walk up to the door, someone knocks on it.	You're sitting upstairs when someone knocks at the front door.
Right next to you, a machine gun is firing.	In the distance, a machine gun is firing.
The sheep walks up to you and bleats.	The sheep wanders to the other side of the hill from you and bleats.
As you hold the frog in your hands, it starts to croak.	At the other end of the pond, a frog starts to croak.
You stand in front of the toilet and flush it.	Someone upstairs flushes the toilet.
You stand next to the waterfall as the water cascades down.	You stand across the valley from the waterfall, as the water cascades down.
You quickly open the can of soda.	Across the bar, a man quickly opens a can of soda.
As you walk through the forest, branches crack under your feet.	Somewhere off in the forest, branches are cracking under someone's feet.

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

References

Barnett, V. & T. Lewis. 1978. *Outliers in statistical data*. New York: Wiley.

Barsalou, L. W. 2002. Being there conceptually: Simulating categories in preparation for situated action. In N. L. Stein, P. J. Bauer & M. Rabinowitz (eds.), *Representation, memory, and development: Essays in honor of Jean Mandler*, 1–15. Mahwah, NJ: Erlbaum.

Barsalou, L. W. 2008. Grounded cognition. *Annual Review of Psychology* 59. 617–645.

Barsalou, L. W. 2009. Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364. 1281–1289.

Barsalou, L. W., A. Santos, W. K. Simmons & C. D. Wilson. 2008. Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg & A. C. Graesser (eds.), *Symbols, embodiment, and meaning*, 245–283. Oxford: Oxford University Press.

Bergen, B. 2007. Experimental methods for simulation semantics. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson & M. J. Spivey (eds.), *Methods in cognitive linguistics*, 277–301. Amsterdam: John Benjamins.

Bergen, B., S. Lindsay, T. Matlock & S. Narayanan. 2007. Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science* 31. 733–764.

Bergen, B. & K. Wheeler. 2005. Sentence understanding engages motor processes. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*, 238–243.

Bergen, B. & K. Wheeler. 2010. Grammatical aspect and mental simulation. *Brain and Language* 112(3). 150–158.

Coleman, P. D. 1968. Dual role of frequency spectrum in determination of auditory distance. *Journal of the Acoustical Society of America* 44(2). 631–632.

Boersma, P. & D. Weenink. 2009. *Praat: Doing phonetics by computer* (Version 5.1.05) [Computer program]. Retrieved March 9, 2010, from http://www.praat.org/.

Borghi, A. M., A. M. Glenberg & M. P. Kaschak. 2004. Putting words in perspective. *Memory and Cognition* 32. 863–873.

Brunyé, T. T., T. Ditman, C. R. Mahoney, J. S. Augustyn & H. A. Taylor. 2009. When you and I share perspectives: Pronouns modulate perspective-taking during narrative comprehension. *Psychological Science* 20. 27–32.

van Danzig, S., D. Pecher, R. Zeelenberg & L. W. Barsalou. 2008. Perceptual processing affects conceptual processing. *Cognitive Science* 32. 579–590.

Ditman, T., T. T. Brunyé, C. R. Mahoney & H. A. Taylor. 2010. Simulating an enactment effect: Pronouns guide action simulation during narrative comprehension. *Cognition* 115. 172–178.

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

Glenberg, A. M. & D. A. Robertson. 2000. Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language* 43(3). 379–401.

Haber, R. N. & C. A. Levin. 2001. The independence of size perception and distance perception. *Perception and Psychophysics* 63(7). 1140–1152.

Horton, W. S. & D. N. Rapp. 2003. Out of sight, out of mind: Occlusion and the accesibility of information in narrative comprehension. *Psychonomic Bulletin and Review* 10(1). 104–110.

Ingard, U. 1953. A review of the influence of meteorological conditions on sound propagation. *Journal of the Acoustical Society of America* 25. 405–411.

Kaschak, M. P., C. J. Madden, D. J. Therriault, R. H. Yaxley, M. E. Aveyard, A. A. Blanchard & R. A. Zwaan. 2005. Perception of motion affects language processing. *Cognition* 94. B79–B89.

Kaschak, M. P., R. A. Zwaan, M. E. Aveyard & R. H. Yaxley. 2006. Perception of auditory motion affects language processing. *Cognitive Science* 30. 733–744.

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press

Kosslyn, S. M., G. Ganis & W. L. Thompson. 2001. Neural foundations of imagery. *Nature Neuroscience Reviews* 2. 635–642.
Mahon, B. Z. & A. Caramazza. 2008. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology - Paris* 102. 59–70.
Morrow, D. G., S. L. Greenspan & G. H. Bower. 1987. Accessibility and situation models in narrative comprehension. *Journal of Memory and Language* 26. 165–187.
Pecher, D., S. van Dantzig, R. A. Zwaan & R. Zeelenberg, R. 2009. Language comprehenders retain implied shape and orientation of objects. *The Quarterly Journal of Experimental Psychology* 62. 1108–1114.
Schneider, W., A. Eschman & A. Zuccolotto. 2002. *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools, Inc.
Sereno, S. C., P. J. O'Donnell & M. E. Sereno. 2009. Size matters: Bigger is faster. *The Quarterly Journal of Experimental Psychology* 62(6). 1115–1122.
Stanfield, R. A. & R. A. Zwaan. 2001. The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science* 12. 153–156.
Sugg, M. J. & J. Polich. 1995. P300 from auditory stimuli: intensity and frequency effects. *Journal of Biological Psychology* 41(3). 255–269.
Taylor, L. J. & R. A. Zwaan. 2009. Action in cognition: The case of language. *Language and Cognition* 1. 45–58.
Vermeulen, N., O. Corneille & P. M. Niedenthal. 2008. Sensory load incurs conceptual processing costs. *Cognition* 109. 287–294.
Vuilleumier, P., R. N. Henson, J. Driver & R. J. Dolan. 2002. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience* 5. 491–499.
Wheeler, K. & B. Bergen. 2010. Meaning in the palm of your hand. In S. Rice & J. Newman (eds.), *Empirical and experimental methods in conceptual structure, discourse, and language*, 150–158. Stanford: CSLI.
Woodworth, R. S. & H. Schlosberg. 1954. *Experimental psychology*. New York: Holt.
Yaxley, R. H. & R. A. Zwaan. 2007. Simulating visibility during language comprehension. *Cognition* 105(1). 229–236.
Zahorik, P. 2002. Assessing auditory distance perception using virtual acoustics. *Journal of the Acoustical Society of America* 111(4). 1832–1846.
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), 35–62. New York: Academic Press.
Zwaan, R. A. & G. A. Radvansky. 1998. Situation models in language comprehension and memory. *Psychological Bulletin* 123(2). 162–185.
Zwaan, R. A., R. A. Stanfield & R. H. Yaxley. 2002. Language comprehenders mentally represent the shapes of objects. *Psychological Science* 13. 168–171.

https://doi.org/10.1515/langcog-2012-0001 Published online by Cambridge University Press