Representing spatial shifts in event processing

Glenn Patrick Williams (g.w.williams@dundee.ac.uk)

Anuenue Kukona (a.b.bakerkukona@dundee.ac.uk)

Yuki Kamide (y.kamide@dundee.ac.uk)

School of Psychology, University of Dundee,
DD1 4HN, UK

Abstract

Two experiments explored the effects of changes in distance and location on the accessibility of event-related information during language comprehension. In Experiment 1, listeners viewed visual scenes depicting a location containing several objects, while they listened to narratives describing an agent either staying in that initial location, or moving to a new one (either close or far away), and then thinking about one of the depicted objects. We found that eye movements to these objects were modulated (reduced) by changes in location, rather than distance. In Experiment 2, listeners viewed scenes depicting two rooms, while they listened to narratives describing an object moving either between the rooms, or within one room. When the object was mentioned following the event, we found fewer eye movements to it when the movement occurred between rooms. We discuss these results in relation to the Event Horizon model.

Keywords: Event cognition; Mental models; Situation models; Spatial processing; Motion events; Visual world paradigm.

Introduction

When understanding a narrative, we must track information along a number of (often) changing dimensions. For example, *what* is being referred to, and *where* is it? Language comprehenders are assumed to activate a set of mental representations that contain information needed to establish a coherent situation (or mental) model of the events described in a narrative (Glenberg, Meyer, & Lindem, 1987). But what is maintained in a situation model, and how does the structure of a situation model affect comprehension?

Location representation

A series of experiments by Bower and colleagues (Morrow, Greenspan, & Bower, 1987; Rinck & Bower, 1995, 2000) aimed to address whether spatial information is retained in a situation model. They found that language comprehenders often took the perspective of an agent, and focused on information associated with the agent's current location, while suppressing information associated with other locations. Critically, these experiments reported a *spatial gradient of accessibility* (Rinck & Bower, 1995), in which the further the agent travelled from previous locations, the harder it was for comprehenders to retrieve information associated with these locations. However, this effect was subsequently shown to be due to the number of 'event boundaries' encountered, rather than the distance travelled. When controlling for distance, but manipulating the number of *event boundaries* – specifically, doorways – crossed, accessibility for information decreased (Radvansky & Copeland, 2006; Rinck, Hähnel, Bower, & Glowalla, 1997), suggesting that "walking through doorways causes forgetting" (Radvansky & Copeland: p.1154).

Competing representations

Why does crossing an event boundary affect accessibility? The Event Horizon model (Radvansky, 2012) claims that information is structured around, and segmented according to, 'event boundaries'. If a boundary is crossed, information either side of it is segmented into separate events. This account also argues that (i) one event will be more activated or more in focus than others in comprehenders' working memory; and (ii) when objects are contained in multiple events, competition can occur (Radvansky & Zacks, 2011). As such, this model predicts a cost to switching focus from one event to another, and for accessing a single representation of an object that is represented in two events.

Current experiments

Although the current evidence suggests that situation models are structured around event boundaries, and not distance, much of this evidence depends on explicit memory-based tasks, which often take place long after the processing of critical linguistic input. Indeed, both distance and event boundaries have been shown to influence the accessibility of information under certain task conditions (Rinck & Denis, 2004). Here, we addressed two issues: in the absence of an overt task, (i) are situation models organised around distance or event boundaries (**Experiment 1**); and (ii) do event-based effects stem from event switching costs, or memory load costs (i.e., of maintaining multiple events; **Experiment 2**)? We used the visual-world eye-tracking paradigm (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to explore these issues.

Experiment 1

Participants viewed visual scenes such as Figure 1, while listening to narratives such as:

(1) "The boy gazes at the picture and the chair in the room. An hour later, he sings in the {(a) room /(b) kitchen /(c) playground}, and thinks about how the picture is very beautiful."

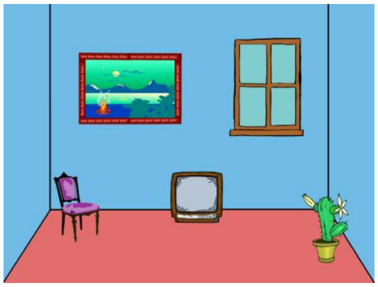

Figure 1: Example visual scene used in Experiment 1

While the agent (boy) remains in the initial location in (1a), he moves to a near location in (1b), and a far location in (1c). Here, we use the terms 'event' and 'event boundary' similarly to previous studies (e.g., Radvansky, 2012): an 'event' consists of a series of actions and states (gazing the picture, gazing the chair etc). These actions/states are considered as one event unless there is an 'event boundary' (e.g., doorway) between them. In our examples, we assume that there is only one event in (1a), whereas there are two in (1b) and (1c), due to the event boundary between the original and new locations.

By examining fixations to the target *picture* during the discourse-final "picture", and the target *picture* and competitor *chair* in the window preceding the discourse-final "picture", we can explore whether distance or event boundaries affect accessibility. Both distance (e.g., Rinck & Bower, 1995) and event boundary hypotheses (e.g., Event Horizon model; Radvansky, 2012) predict more fixations to the target in (1a) than (1b) and (1c). However, the distance account predicts more looks to the target in (1b) than (1c), while the event boundary hypothesis predicts no difference.

Methods

Participants Fifty-five native English speakers from the University of Dundee participated for course credit or £5.

Materials We created thirty-six items (1) with same, near, and far forms (based on preliminary norming of the distances between locations). Only one form was presented to each participant. A temporal phrase (e.g., 'An hour later') was added to the second sentence of the discourse to keep the duration of the time lapse constant across conditions. Half of the items had the first-mentioned object ("picture") as the discourse-final noun, and the other half had the second-mentioned object ("chair") as the discourse-final noun. In an additional thirty-six filler items, the discourse-final noun was not referred to earlier in the discourse.

Procedure We used an SR Research EyeLink-II head-mounted eye-tracker and a 'look and listen' task. Participants were instructed to simply look at the visual stimulus and listen to the auditory sentences. We used the 'blank screen' paradigm (cf., Altmann, 2004), in which scenes on each trial were first presented alone, and then replaced with a 'blank' screen, and the accompanying auditory stimuli. Each scene was displayed for 7,000 ms, followed by a 1,000 ms preview of the blank screen (mean trial duration = 13,264 ms). Each trial was terminated 4,000 ms after the offset of the spoken stimulus. The experiment lasted approximately 45 minutes.

Results

Eye movements were analysed during two time windows: the discourse-final noun ("picture": mean duration = 570ms) and the window preceding the discourse-final noun ("about how the": mean duration = 580ms). The latter was included to test for anticipation of the discourse-final noun: in half of the experimental items, the target object was the first-mentioned object in the first sentence ("picture"), whereas in the other half it was the second-mentioned object ("chair"), and thus we averaged across these objects in the analysis. The mean proportions of fixations to these objects are presented in Figure 2.

Proportions of fixations by participants and items were submitted to multilevel linear mixed-effects analyses with an empirical logit transformation (see Barr, 2008).

Three separate planned comparisons were performed for each time window: room × kitchen, room × playground, and kitchen × playground. *P*-values were computed using a model comparisons approach (change in log likelihood).

During "picture" Model comparisons for the mean proportions of fixations to the target *picture* during the discourse-final noun ("picture") revealed no significant difference between the three conditions; all ps > .05.

During "about how the" Model comparisons for the mean proportions of fixations to the target *picture* and competitor *chair* during "about how the" revealed that fixations to these objects were reliably greater in the 'room' condition when compared to the 'kitchen' condition (ps < .05). Fixations to the *picture* and *chair* were also greater in the 'room' condition when compared to the 'playground' condition by items ($p < .05$), and marginally by participants ($p < .10$). No significant differences were found between the proportions of fixations to the *picture* and *chair* between the 'kitchen' and 'playground' ($p > .05$) conditions.

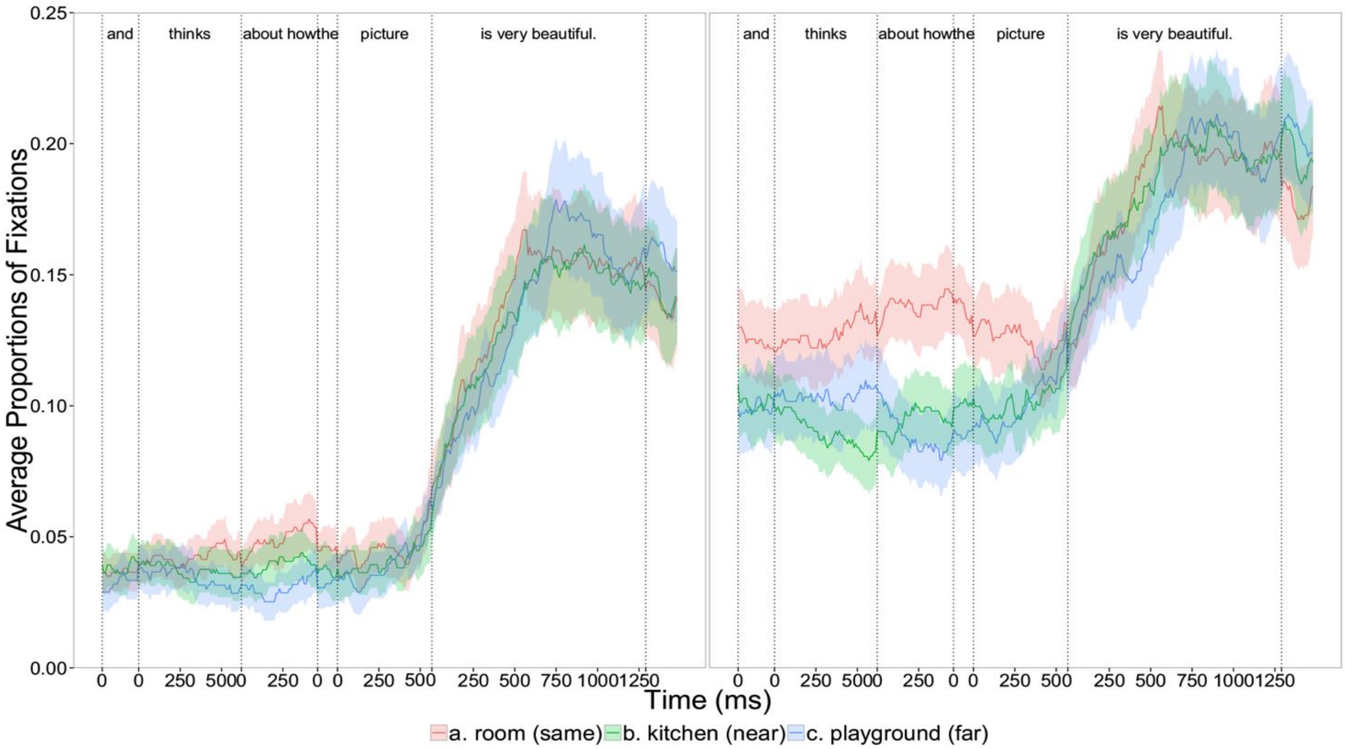

Figure 2. Mean proportions of fixations to the target *picture* (left) and to the target *picture* and competitor *chair* (right) in Experiment 1 (error bars show SE).

Discussion

During the discourse-final noun ("picture") we found no differences in fixations to the target *picture* across conditions. However, during "about how the", listeners were more likely to fixate both the target *picture* and the other object referred to, the competitor *chair*, in the 'room' (same) condition compared to both the 'kitchen' (near) and 'playground' (far) conditions. Consistent with the Event Horizon model (Radvansky, 2012), reactivation of potential targets (prior to hearing the discourse-final noun) was modulated by the number of event boundaries crossed, and not by distance (e.g., between the boy and *picture*/*chair*). Critically, these effects were observed during online language comprehension (in fact, before referents were re-mentioned), without a memory-based task.

Although no effect of condition on accessibility was found during the discourse-final noun, this may be because these eye movements were guided primarily by bottom-up referential processing (i.e., hearing "picture" triggered looking to the depicted *picture*).

Thus, the results of Experiment 1 are compatible with the event boundary, rather than distance, account. However, there are two other accounts of our results. First, in the 'room' condition, the target *picture* was spatially associated with the agent in his/her final location. Since previous studies have found that objects that are spatially close to the agent tend to be 'foregrounded' in discourse processing (e.g., Glenberg et al., 1987), it is possible that access to the picture was easier in the 'room' condition due to the foregrounding of this object, compared with the other conditions in which it was 'backgrounded'. Second, during the discourse-final noun the visual scene corresponded to the agent's location in the 'room' condition, but not in the other conditions. Thus, there was a match between the agents' final location and the visual scene in the 'room' condition, but a mismatch between his/her final location and the visual scene in the other conditions, which may also have influenced accessibility.

Experiment 2 aimed to address the locus of the 'event boundary' effects observed in Experiment 1. In Experiment 2, similar to Experiment 1, we compared discourses in which an event boundary (doorway) was either crossed or not. However, unlike Experiment 1, the critical referent object remained with the agent in both 'boundary' and 'no boundary' conditions, allowing the referent to be equally foregrounded in both conditions. Additionally, visual scenes depicted both original and new locations. Finally, we kept the distance between original and new locations constant across conditions. Thus, the presence or absence of an event boundary was the only difference between conditions.

Experiment 2

Altmann and Kamide (2009) showed that when an object in the 'visual world' is described as moving to a new location, listeners fixate its new location more than when it does not move. The current experiment explored whether the introduction of an event boundary along the object's path influences accessibility of that object. Participants were presented with one of two visual scenes (Figure 3a or 3b), while listening to sentences such as:

(2) "The woman will take the book to the table. Then she will study the painting and pick up the book."

Figure 3: Example visual scenes used in Experiment 2: (a) Boundary condition (left) and (b) No-boundary condition (right).

In the Boundary condition (Figure 3a) a *doorway* separating the rooms was located between the critical object (*book*) and the goal (*table*), whereas in the No-boundary condition (Figure 3b) the critical object and goal were in the same room. Thus, during the narrative the critical object moves across an event boundary in the Boundary condition, but not in the No-boundary condition.

Critically, in the current experiment the discourse-final noun referred to the goal location that the book was moved to after the movement event in both conditions. However, in the No-boundary condition, both representations of the book were within one event, whereas in the Boundary condition they were in separate events. If the event boundary effects reported in Experiment 1 are due to (linguistic and/or visual) foregrounding, then we expect no differences across conditions. However, if our effects are due to an event boundary, then we expect one of two outcomes. One possibility is that event boundaries make object representations that are not in the event in focus less accessible. Thus, during the discourse-final noun (e.g., "book") we expect fewer looks to the *book* in the Boundary condition than No-boundary condition (and more looks to the *table* in the Boundary condition than No-boundary condition: i.e., *book*: No-boundary > Boundary; *table*: No-boundary < Boundary). An alternative possibility is that when objects occur in multiple events, these representations compete, and reduce accessibility to both representations (i.e., whether they are in the event in focus or not; e.g., Radvansky & Coupland, 2006). As discussed in the Introduction, the Event Horizon model (Radvansky, 2012) takes this position, arguing for increased competition and costs between representations in different events. Thus, following the model, we expect fewer looks to both the *book* and *table* in the Boundary condition than the No-Boundary condition (i.e., *book*: No-Boundary > Boundary; *table*: No-Boundary > Boundary).

Methods

Participants Thirty native English speakers from the University of Dundee participated for course credit or £5.

Materials We created twenty pairs of experimental pictures (e.g., Figure 3a & 3b), which depicted two rooms separated by a doorway. The distance between the critical object and goal were identical across pairs. The direction of movement and orientation of the boundaries was also counterbalanced across items. In an additional twenty filler items, the critical object was also not located in the centre of the scene, and in half of these, the discourse-final noun did not refer to a moved object.

Procedure Again, we used a 'look and listen' task. However, unlike the 'blank screen' paradigm used in Experiment 1, visual scenes and auditory stimuli were presented concurrently, in order to ensure that participants had visual information about the spatial relations among objects (e.g., *book*, *doorway*, *table*) when they were processing the linguistic discourse. Each scene was displayed for 1,000 ms, after which the critical discourse played over speakers. The visual scene remained on screen for the duration of the trial (mean trial duration = 5,787 ms). Each trial was terminated 4,000 ms after the offset of the spoken stimulus. The experiment lasted approximately 25 minutes.

Results

Eye movements were analysed during the discourse-final "book". The regions of interest were the critical object (*book*) and the goal (*table*). The mean proportions of fixations to these objects are presented in Figure 4. Again, proportions of fixations were submitted to empirical logit analyses. Two separate planned comparisons were performed: Boundary vs. No-boundary conditions for both critical objects and goals. Unlike in Experiment 1, we do not report results for the window preceding the discourse-final noun because of the greater number of linguistically-focused visual objects, and the greater complexity of the fillers, which made anticipatory effects less likely.

Critical object (*book*) Model comparisons revealed no significant difference in the proportions of fixations towards the critical *book* between the Boundary and No-boundary conditions ($ps > .05$) during the discourse-final "book".

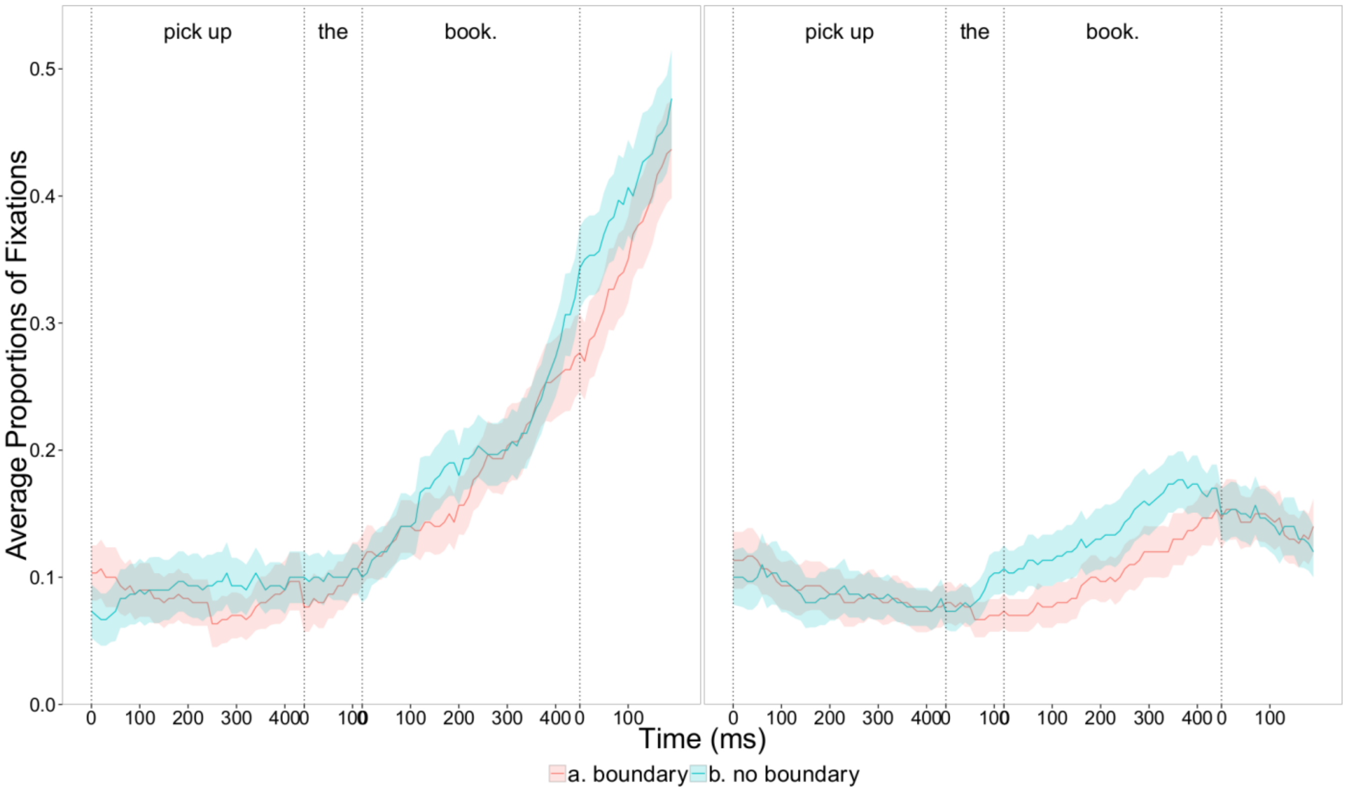

Figure 4. Mean proportions of fixations to the critical object (*book:* left) and goal (*table*: right) in Experiment 2 (error bars show SE).

Goal (*table*) Model comparisons revealed a significantly higher proportion of fixations towards the goal (*table*) during the discourse-final noun ("book") in the No-boundary condition than the Boundary condition ($ps < .05$).

Discussion

During the discourse-final noun ("book": mean duration = 530ms) we found no difference in fixations to the critical object (*book*) between the Boundary and No-boundary conditions. However, we found that listeners were more likely to fixate the goal (*table*) in the No-boundary condition than the Boundary condition. Although we predicted that differences in fixations to the goal would be accompanied by differences in fixations to the critical object (*book*), this mismatch could be due to a number of other factors. For example, because the visual scene was presented concurrently in the current experiment, it is plausible that bottom-up referential processing drove fixations to objects (e.g., book) when they were mentioned, obscuring differences between the conditions. By contrast, fixations to the goal depended on listeners' situation models of the narrative, because critical objects were never depicted there. Thus, fixations to the goal may provide better insight into comprehenders' mental representations than fixations to the critical object (*book*). Additionally, a follow-up study is currently underway to address this issue, which utilizes the blank screen paradigm. By removing the visual scene prior to the onset of the narrative, it is expected that bottom-up referentially-driven fixations to depicted objects will be reduced, and that comprehenders will rely to a greater degree on their mental representations.

Further, another account of the current results is that our fixation patterns were driven by the durations, or the complexities, of the activities involved. For example, the Boundary condition presumably requires agents to open the door when moving the critical object, which both increases the activity's duration, and adds to its complexity (Coll-Florit & Gennari, 2011). We are currently conducting an experiment in which these confounds will be minimised by: (i) replacing these closed door event boundaries with open arches; and (ii) norming the duration and complexity of our described events.

Finally, the results of the current experiment are partially compatible with the Event Horizon model, as reflected in the increased difficulty of accessing the goal when there is an event boundary (vs. not).

General Discussion

In two experiments, we found evidence that event boundaries modulated accessibility for information during online language comprehension. Consistent with Radvansky (2012), Experiment 1 showed that comprehenders anticipatorily activated potential discourse referents based on the location of an agent relative to these referents within a narrative: reactivating referents across an event boundary (prior to hearing the discourse-final noun) was more difficult than reactivating referents within the same event.

Experiment 2 explored whether this effect was driven primarily by greater accessibility for items in the same location as the agent, or by the presence of an event boundary. The results of Experiment 2 indicated that competition between the representations of an object across an event boundary is likely to drive the difficulty of accessing information about that object.

Although these findings suggest that representing an object across an event boundary generates competition, this does not rule out the possibility that foregrounding played a contributing factor in the first experiment. Yet, these two experiments show that comprehenders spontaneously form and update spatial situation models during spoken language processing, even without an overt task, and that these models guide fixations and comprehension. Crucially, they reflect information about both: (i) which object will be mentioned (Experiment 1); and (ii) which representation of an object is relevant (Experiment 2).

Finally, parallels can be drawn between the findings of Experiment 2 and those of Hindy, Altmann, Kalenik, and Thompson-Schill (2012). Hindy et al. report that conflict-associated brain regions are activated during sentences that describe object state-changes (e.g., “The squirrel will *crack* the acorn”, [substantial change] vs. “The squirrel will *sniff* the acorn” [minimal/no change]). These results have been taken to indicate that when an object’s pre- and post-event states differ substantially due to a substantial (“crack”), rather than minimal (“sniff”), change, greater conflict is generated due to the difficulty of supressing the more dissimilar competing representations (i.e., dissimilarity-based interference). This could offer a potential insight into why and how conflict occurs when accessing a specific representation of an object after a movement event: When crossing an event boundary, each representation of an object will be more dissimilar than if no boundary was crossed because those representations will be associated with different events. Thus, like Hindy et al, there will be greater (dissimilarity-based) interference among the more dissimilar competing representations.

Overall, our results suggest that during online discourse processing, listeners track and segment information according to event boundaries. Moreover, it is shown that the mechanism of the mapping between two entities (agent in the current location and critical object) or the competition between two instantiations of the same object (book on the floor vs. book on the table) is largely influenced by event boundaries during incremental establishment of a coherent situation model.

References

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: the “blank screen paradigm”. *Cognition, 93*(2), 79–87.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: eye movements and mental representation. *Cognition, 111*(1), 55–71.

Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4), 457–474.

Bates, D. (2007). *lme4: Linear mixed-effects models using S4 classes*. (R package version 0.99875-2).

Coll-Florit, M. & Gennari, S. (2011). Time in language: event duration in language comprehension. *Cognitive Psychology, 62*(1), 41–79.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text Comprehension . *Journal of Memory and Language*, *26*(1), 69–83.

Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object state-changes on event processing: do objects compete with themselves? *The Journal of Neuroscience, 32*(17), 5795–803.

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language, 26*, 165–187.

Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science, 21*(4), 269–272.

Radvansky, G., & Zacks, J. M. (2011). Event Perception. *WIREs Cognitive Science, 2*(6), 608–620.

Radvansky, Gabriel A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: situation models and experienced space. *Memory & Cognition, 34*(5), 1150–1156.

Radvansky, Gabriel A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *Quarterly Journal of Experimental Psychology, 64*(8), 1632–1645.

Rinck, M., & Bower, G. H. (1995). Anaphora resolution and the focus of attention in situation models. *Journal of Memory and Language, 1*(34), 110–131.

Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. *Memory & Cognition, 28*(8), 1310–1320.

Rinck, M., & Denis, M. (2004). The metrics of spatial distance traversed during mental imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1211–1218.

Rinck, M., Hähnel, A., Bower, G. H., & Glowalla, U. (1997). The metrics of spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 622–637.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.