

Implied motion language can influence visual spatial memory

David. W. Vinson¹ · Jan Engelen² · Rolf A. Zwaan² · Teenie Matlock¹ · Rick Dale¹

Published online: 15 March 2017
© Psychonomic Society, Inc. 2017

Abstract How do language and vision interact? Specifically, what impact can language have on visual processing, especially related to spatial memory? What are typically considered errors in visual processing, such as remembering the location of an object to be farther along its motion trajectory than it actually is, can be explained as perceptual achievements that are driven by our ability to anticipate future events. In two experiments, we tested whether the prior presentation of motion language influences visual spatial memory in ways that afford greater perceptual prediction. Experiment 1 showed that motion language influenced judgments for the spatial memory of an object beyond the known effects of implied motion present in the image itself. Experiment 2 replicated this finding. Our findings support a theory of perception as prediction.

Keywords Implied motion · Motion language · Spatial cognition · Visual memory · Prediction

Imagine that you are driving down a highway and see a car start to drift into your lane. To avoid a head-on collision, you swerve to your right. Dodging the car in this way requires you to anticipate its future position relative to your own. In a matter of seconds, or even less, you are able to do this anticipation and avoid a collision. Dodging cars and similar everyday actions involves the execution of motor commands

that rely on the accuracy of their predicted consequences (Flanagan, Vetter, Johansson, & Wolpert, 2003). The strength of the relationship between motor actions and their predicted outcomes has inspired the view that the brain is a “predictive machine” (Bar, 2003; Bar et al., 2006; Barsalou, 1999; Barto, Sutton, & Watkins, 1990; Bubic, von Cramon, & Schubotz, 2010; Clark, 2007, 2013a, b; Elman & McClelland, 1988; Hommel, Müsseler, Aschersleben, & Prinz, 2001; Jordan, 2013; Kinsbourne & Jordan, 2009; Körding & König, 2000; Neisser, 1976; Schubotz, 2007; Spratling, 2010; Ullman, 1995).

In this general approach, cognition is not just about accurately *representing* the size, shape, and position of various objects in one’s visual array, but *predicting* those aspects with respect to possible future actions (Gibson, 1979; see Morsella, Godwin, Jantz, Krieger, & Gazzaley, 2015, for a review), guided by past experiences (Bar, 2009; Barsalou, 1999, 2009), in evolutionarily adaptive ways (Gibson, 1979; Proffitt, 2006; Witt, 2011). According to this view, the constant stream of visual information is influenced by a variety of prior *top-down* information (Goldstone, de Leeuw, & Landy, 2015; Kveraga, Ghuman, & Bar, 2007; Lamme & Roelfsema, 2000; Lupyan & Clark, 2015; Stins & van Leeuwen, 1993; Sugase, Yamane, Ueno, & Kawano, 1999; Vinson et al., 2016), including conceptual and categorical labeling (Lupyan, 2012; Lupyan, Thompson-Schill, & Swingley, 2010), language (Skipper, 2014; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), goals (Büchel & Friston, 1997), and past actions (e.g., memories: Friston, 2005; Hummel & Holyoak, 2003; Jones, Curran, Mozer, & Wilder, 2013; Jordan & Hunsinger, 2008).

If prediction is ubiquitous in everyday cognition, then it should operate in many domains, not just in simple perception. Our *memory* for a scene or event should also be influenced by the dynamics of prediction. In the present study, we examined the influence of top-down information on visual spatial

✉ David. W. Vinson
dvinson@ucmerced.edu

¹ University of California, 5200 N. Lake Rd., Merced, CA 95343, USA

² Erasmus University of Rotterdam, Rotterdam, The Netherlands

memory. Specifically, we tested whether the remembered spatial location of a car positioned on a hill was influenced by the prior presentation of motion language. Cuing participants to probable future changes in a scene may affect spatial memory. If so, this effect should be consistent with what is predicted about the scene and how its objects may move. For example, memory for the location of the car may be displaced, consistent with its potential movement under gravity. If this effect obtains, then spatial memory even for static scenes is influenced by top-down information that implies future motion.

Such an effect may not reflect “flaws” in our ability to remember spatial locations. It may be reasonable to think of mislocations in spatial memory as a kind of positive cognitive outcome (Vinson, Jordan, & Hund, 2017): Our memory is aligning with expectations from other sources. For example, when phrasing a question about a car accident, use of the word “smashed” (vs. “bumped”) will prompt participants to incorrectly recall glass when none was actually involved (Loftus & Palmer, 1974). In classic attentional blindness, we have an inclination *not* to perceive objects irrelevant to a current goal, such as a gorilla casually walking around when we are trying to count the number of passes made by basketball players (Simons & Chabris, 1999; see also Drew, Vö, & Wolfe, 2013). Prior stimuli can also impact the next judgments. Participants tend to remember features of faces as being more similar to those previously seen (Liberman, Fischer, & Whitney, 2014). These various “mistakes” may simply be the result of our cognitive system “settling in on a hypothesis that maximizes the posterior probability of observed sensory data” (see Clark, 2013a). Put another way, errors in judgment may be a reflection of the cognitive system’s adaptation to the statistical regularities in its environment (Qian & Aslin, 2014). In the present case, we expected that spatial memory for an object can be displaced predictably by how the object is implied to be moving in the scene. In our task, the car did *not* move in the static scene, but language in the instructions might guide spatial memory in the direction of likely movement.

Top-down conceptual information

Some research in the broad area of embodied language and cognition has shown that information from multiple modalities collectively influences response latencies to stimuli (Glenberg et al., 2010; Madden & Zwaan, 2003; Santos, Chaigneau, Simmons, & Barsalou, 2011; Stanfield & Zwaan, 2001; for replications, see Zwaan & Pecher, 2012). Reading a sentence that depicts the orientation of an object facilitates or inhibits visual object recognition depending on whether the object’s orientation matches or mismatches, respectively (Zwaan, Stanfield, & Yaxley, 2002). Such findings have been used to support claims that we simulate or create mental models of actions and situations that unfold over time (Barsalou, 1999;

Matlock, 2004; Morrow & Clark, 1988; Radvansky, Zwaan, Federico, & Franklin, 1998; Zwaan & Radvansky, 1998). One possibility is that prior information may encourage or discourage simulations from unfolding during visual processing. For instance, humans are more likely to perceive visual motion at a perceptual threshold when they are first presented with matching linguistic information (e.g., up, down, left, or right; Meteyard, Bahrami, & Vigliocco, 2007). That is, language may benefit the listener by mediating the effects of motion presented in vision (Gennari, Sloman, Malt, & Fitch, 2002). Further support for this hypothesis comes from Coventry and colleagues (Coventry, Christophel, Fehr, Valdés-Conroy, & Herrmann, 2013). They found that when spatial prepositions match the observed relationship between two static images that could imply motion (e.g., a vertical preposition like *over* with an image of a bag of pasta oriented as if it were starting to pour its contents into a pot positioned below it), there was greater cortical activation in areas related to motion perception than when presented with the images alone.

Moreover, conceptual knowledge presented together with visual information may influence simulations in visual perception. When presented with a sequence of rocket-like static images that imply upward motion, people remember the rockets’ locations as being farther displaced in the direction congruent with motion than when presented with steeple-like images (Reed & Vinson, 1996). Spatial memory is further influenced by the simultaneous presentation of information relevant to the stimuli’s possible motion (Hubbard, 2005, 2014), such as implied friction (Jordan & Hunsinger, 2008), implied velocity (Freyd & Finke, 1984), implied gravity (De Sá Teixeira, Hecht, & Oliveira, 2013; Hubbard & Bharucha, 1988), and stimulus orientation (Vinson, Abney, Dale, & Matlock, 2014). Here again, these findings are thought to be the result of mentally simulating an object’s motion (Freyd, 1987; see also Hubbard, 1995).

Motion information in language and vision may be integrated into the same perceptual experience. If so, their joint effects should be detectable via postperceptual judgments. We predict that motion language will significantly affect participants’ spatial memory in one of two ways: (1) Motion language will mediate known effects of implied visual motion on spatial memory. When the motion language is congruent with an object’s implied motion presented visually, memory for the object’s location will be displaced in the direction of implied motion. However, the amount of displacement will be no greater than if only presented with visual implied motion. Furthermore, language that is incongruent with implied motion in vision will result in no effect of visual motion on spatial memory. That is, language acts as a mediator for visual implied motion. Alternatively, (2) motion language will encourage motion simulations resulting in greater spatial memory displacement. Linguistic motion may provide novel information about the stimulus in question. As a result, the remembered location of an object may be displaced farther along its implied trajectory than can be explained by the effects of

implied motion present in vision alone. If so, when the language is congruent with visual implied motion, the remembered location of the object will be displaced farther along its implied motion trajectory than when the object is presented only with visual implied motion. Moreover, when the language is incongruent, memory for the object's location may appear to be unaffected, but this would be due to simultaneous incongruent motion simulations that cancel out each other's effects.

Present study

In two experiments, we determined how spatial memory is influenced by implied motion in language and vision. Both experiments tested whether congruent or incongruent motion language influenced spatial memory of objects that already imply gravitational motion visually (e.g., a car positioned on a hill). Our present method involves the presentation of a stimulus only once. This is novel in that all other experimental designs in implied motion have consisted of multiple sequentially presented stimuli, which when presented together implied motion (see Hubbard, 2005, 2014, for reviews).¹ In these experiments, the location of the stimulus, when acted upon solely by implied gravity, was remembered as being farther along an implied gravitational trajectory. However, implied vertical motion is also influenced by conceptual knowledge about the stimulus (Reed & Vinson, 1996). If participants anticipate the car's movement to be in the direction the car is facing, then the car orientation will impact its remembered spatial location. However, if the possibility of implied motion that stems from the car's orientation is dependent on contextual information from both the scene itself and linguistic information, then the car may be remembered as farther along a gravitational trajectory, regardless of its orientation, when no language is presented, and farther still when language congruent with implied gravity is presented. When the language is incongruent, the effects of visual implied gravity may not occur.

To assess the effects of language and orientation on car placement, we projected the x -axis onto a gravity dimension (GD) by transforming the x -coordinate plane onto a single dimension aligned to the slope of the hill (see Fig. 1). The gravity dimension, $GD = x/\cos \theta$ — where GD is the hypotenuse of the triangle formed by the x -axis displacement and its right angle — removed the possibility that a significant effect of the participant's placement beyond the car's actual location was simply due to remembering the car as being closer to the middle

of the road (i.e., a possible confound affecting y -axis placement).² That is, we removed the possible confound that moving the car along the y -axis (closer to the center of the road) influenced the participant placements.

We first presented participants with motion language—"The car moved forward," "The car moved backward," or no language—followed by the image of a car facing up or down a hill (see Fig. 2). If there was an effect of visually implied gravitational motion alone, then the car should be remembered as being farther down the hill in the no-language condition. This condition acted as a simple control replicating previous, well-known implied gravitational-motion effects. Both experiments tested whether motion language influenced the spatial memory of an object's location beyond that of visual implied motion.

If language influences spatial memory, either by mediating the effects of visual motion on spatial memory or increasing the richness of motion simulation by providing novel information, then when participants read a sentence that was incongruent with visual implied motion, such as "The car moved *forward*" when the car was facing *uphill*, or "The car moved *backward*" when the car was facing *downhill*, we would find no effect of implied motion on spatial memory. In this sense, the motion language would effectively cancel out any effects of implied motion in vision. This might be the result of the presence of incongruent simulations occurring simultaneously, or of the lack of sufficient congruent simulation, due to the incongruent information from vision and language. If language only mediates the effects of vision on spatial memory, then when motion language was congruent with the implied visual motion, such as "The car moved *forward*" when the car was facing *downhill* or "The car moved *backward*" when the car was facing *uphill*, there should *only* be an effect of implied visual motion, and no additional effect of language. Alternatively, if implied-motion language contributes to richer simulations or predictions of visual information, the car should be remembered as being displaced farther than the effects of implied motion from vision alone would explain. If so, the effect of implied motion in congruent conditions would be greater than the implied motion present from vision alone (i.e., in the no-language conditions).

Experiment 1

Method

A total of 479 Amazon Mechanical Turk³ users from the United States participated in exchange for 0.15 USD. Roughly 75

¹ Freyd (1983) presented participants with a single image of an object followed by a test image. However, the test image, presented 250 ms later, contained that object farther along a motion trajectory. It is unclear that the first image alone was responsible for the observed implied-motion effects, and not the sequential presentation of the same image.

² We thank an anonymous reviewer for thoughtful comments (and a recommendation) on how to avoid this possible confound.

³ Amazon Mechanical Turk is a crowdsourcing marketplace where individuals are compensated for their participation in tasks set up by researchers and companies.

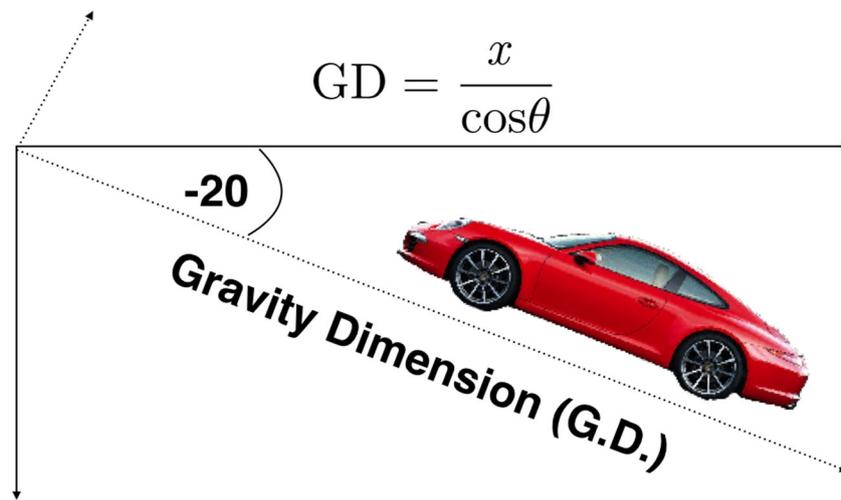


Fig. 1 Coordinate transformation from the original x -coordinate plane onto the gravity dimension. This specific example reflects the gravity dimension and angle from Experiment 1

individuals participated in each of six conditions, replicating the sample sizes from previous research using the same paradigm (Vinson et al., 2014). All participants were required to have an updated version of Adobe Flash player. After choosing to complete the task, participants were directed to a Web link containing an interactive Adobe Flash CS6 program that can be found here: <http://davevinson.com/exp/vzm/2/z2.html>.

Using Adobe Flash CS6, participants observed eight scenes, each with content that differed from the previous scene (see Fig. 2). Scenes were presented relative to the Adobe Flash Width \times Height dimensions (550×400 pixels). The origin of the scene was located at the top left corner of the screen; thus, the x -coordinates were positive, and the y -coordinates, negative. The upper left corner of each object was used as the object's origin in the x -, y -coordinates. The image of the car described below indicates the position of its upper left corner on the screen, unless otherwise specified.

Scene 1 displayed these instructions in black font on a white background: “This is extremely fast! Please pay close attention! On the next screen you will see an image of a car for a split second. After, please indicate where you saw the car by clicking where you think it was. Click anywhere on this screen to begin.” After clicking on the scene, the participant's cursor disappeared and Scene 2 appeared for 2,000 ms, containing a white background with one of the two sentences: “The car moved forward” or “The car moved backward.” A single text box (550×200) with no border contained the sentence (font: Trebuchet, bold; size: 25) presented in the center of the screen, with the top left corner of the box located at $(0, -100)$. This was automatically followed by Scene 3, a black-backdrop mask for a 1,000-ms duration. Scene 4, presented for 1,000 ms, contained the test image of a San Francisco street hill (slope = -0.35) and a red car (275×85) on the street with its origin (upper left corner) located at $(120, -120)$. Scene 5 contained a black-backdrop mask identical to that in Scene 3 in both

presentation and duration. Both black backdrop masks replicated previous experimental masking times used in recent spatial displacement studies (De Sá Teixeira et al., 2013).

Scene 6 was identical to Scene 4 except that the car was omitted from the scene. The cursor reappeared, after having disappeared at the presentation of Scene 2, but this time, instead of a mouse, it was the exact image of the test stimulus from Scene 4 (e.g., the car), oriented in the same way it had been in Scene 4. The precise position of the participant's x -, y -coordinate was the upper left corner—that is, the origin—of the car. Because participants did not see their mouse cursor, no other point of reference was present that could have alluded to where exactly the cursor was located. Participants were to place the car by clicking on the screen where they remembered seeing it last, as had been indicated by the previous instructions. The initial position of the car (i.e., the participant's mouse) was not controlled; wherever the mouse cursor had been previously was where the car appeared. The time it took for participants to decide where they thought they had seen the car was recorded in milliseconds. After, when a final scene appeared, participants were asked to indicate how confident they were in their placement of the car, on a 1–7 Likert-type scale (Scene 7), followed by a short debriefing about the purpose of the study (Scene 8).

It is important to note that cursor-positioning methods are typically used in studies using actual motion (Hubbard, 2005). However, previous research by some of the present authors had successfully used cursor positioning in a static-motion paradigm, further verifying the robust influence of implied-motion images on spatial memory (Vinson et al. 2014).

Results and discussion

A 2 (Orientation: car facing up/down) \times 3 (Language: “The car moved *forward*”/“The car moved *backward*”/no

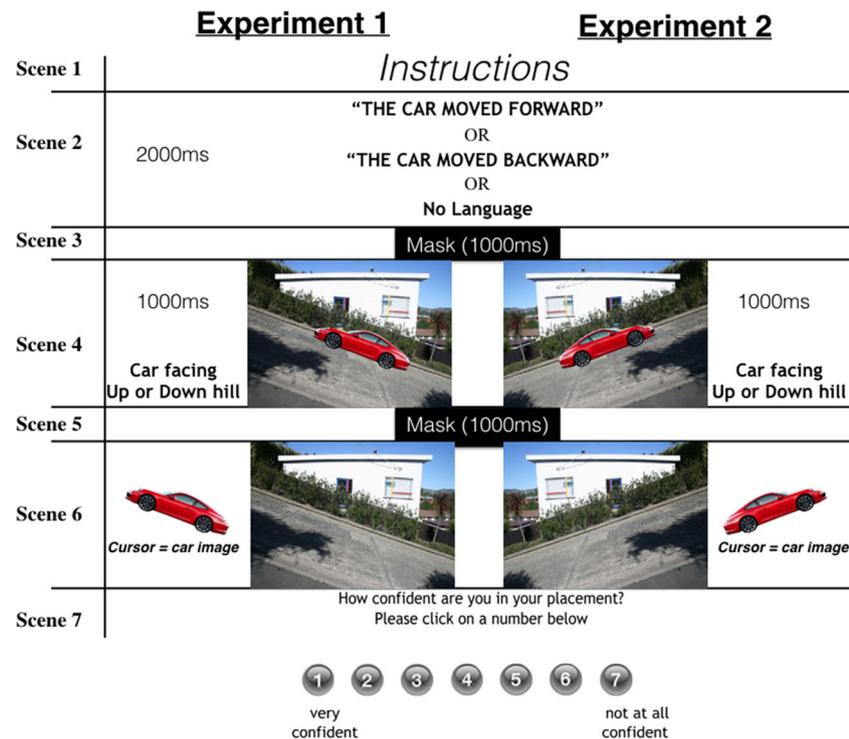


Fig. 2 Design for Experiments 1 and 2. In Scene 2, one of the two sentences was presented for 2,000 ms, followed by a black-screen mask for 1,000 ms. In Scene 4, the image of a car on a hill was presented for 1,000 ms, facing either uphill or downhill, again followed by a black-screen mask for 1,000 ms. In Scene 6, the image of the hill was presented

while the car, once within the scene, was now controlled by the participant’s cursor. Scene 7 was presented after participants had placed the car where they thought it had been and had clicked their mouse. Participants were then asked to rate their confidence in the car placement

language) between-subjects analysis of variance (ANOVA) was performed on placement times, confidence, and GD.

The Gravity Dimension transformation results in a new coordinate for the actual location of the car on ($\mu = 127.7$) and allows for a more direct assessment of the impact of implied gravitational motion on the participant placements. Responses more than three standard deviations away from the mean of the placement times, confidence, and/or GD were removed from all subsequent analyses (a total of 25 participants, ~5%).

Gravity dimension Before analysis, the actual location of the car on GD was subtracted from participant placements centering the data on zero. We found no main effect of language or orientation on GD placements, but there was a significant interaction, $F(1, 448) = 26.76, p < .001, \eta_p^2 = .11$ (Fig. 3). Simple-effects analysis with Tukey HSD correction revealed that in conditions in which language and orientation were congruent with the direction of gravity, the car was displaced significantly farther along in the direction of gravity than in the no-language conditions ($p < .05$) and in conditions in which language and orientation were incongruent with gravity ($p < .001$). Although placements in the no-language conditions were also significantly different from those in one of the incongruent-with-gravity conditions—“backward”–down ($p < .05$)—they were not significantly different from those in “forward”–up.

Spatial memory for the location of the car in conditions in which language and orientation were congruent with gravity was significantly displaced beyond the actual location on the gravity dimension. In both conditions in which language and orientation were incongruent with gravity, the placement location was not significantly different from the actual location of the car [“forward”–up: $M = -2.44, SE = 5.11, t(70) = -0.48$; “backward”–down: $M = -6.24, SE = 3.72, t(80) = -1.68$]. In both conditions in which no language was presented, the car was displaced significantly farther along than its actual location, in the direction of implied gravity [no-language–up: $M = 7.19, SE = 4.56, t(76) = 1.57$; no-language–down: $M = 8.29, SE = 3.75, t(75) = 2.21, p = .03$], replicating well-known effects of implied gravity in visual memory. Finally, in the conditions in which language and orientation were congruent with gravity, the car was displaced significantly farther along in the direction of gravity from the actual location [“backward”–up: $M = 29.03, SE = 5.86, t(70) = 4.96, p < .001$; “forward”–down: $M = 36.30, SE = 6.67, t(77) = 5.45, p < .001$].

Placement times The experimental paradigm was not designed to test placement times. They are reported here solely for the sake of completeness. We found a main effect of language, $F(2, 448) = 5.18, p = .006, \eta_p^2 = .02$, for placement times, such that when language was present, participants were slower to place the car

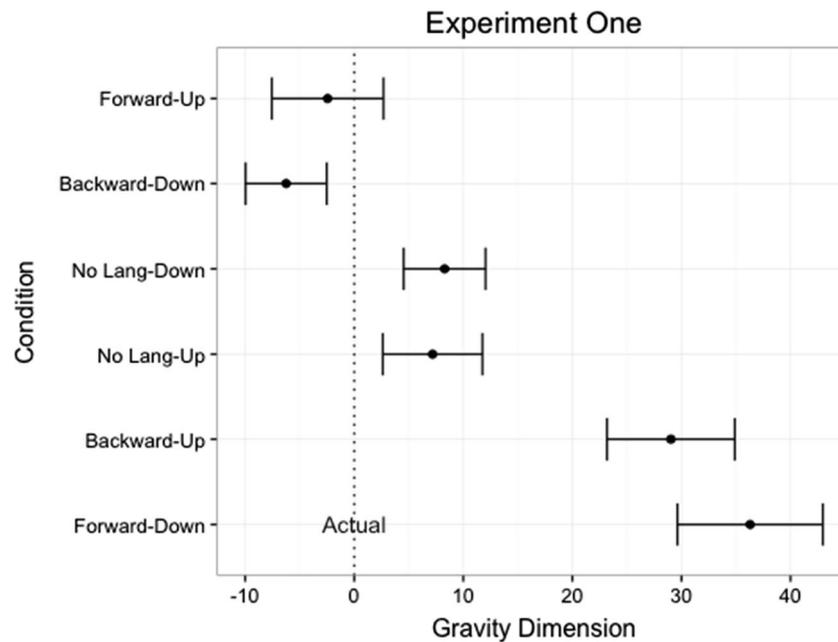


Fig. 3 Means and standard error bars for the remembered location of the car on the gravity dimension, by condition for Experiment 1. The dotted line indicates the actual position of the car, and higher absolute values indicate stronger effects of implied gravitational motion on spatial memory

than when language was not present: no language ($M = 1,912.58$, $SE = 55.53$ ms) < “backward” ($M = 2,105.70$, $SE = 52.32$ ms; $p = .04$) and “forward” ($M = 2,148.83$, $SE = 58.23$ ms; $p = .008$). Although previous research had shown the effects of implied gravitational motion on stimulus placement increases with longer delays between image observation and participant responses (De Sá Teixeira et al., 2013), we did not predict that longer placement times would result in greater displacements, since all participants experienced the same interstimulus interval (1,000 ms). A linear regression analysis in R (lm) with placement times as a predictor of GD placements revealed no main effect of placement times on GD: $F(1, 452) = 0.04$, $p > .05$.

Confidence Neither language nor orientation nor an interaction between the two variables significantly predicted placement confidence ($M = 3.22$, $SE = .09$). Likewise, confidence was not a significant predictor of GD placements: $F(1, 452) < 0.001$, $p > .05$.

Experiment 2

In Experiment 1, the hill ran from left to right. This might have created a confounding variable, in that it matches our reading direction. It is possible that reading direction introduces a rightward bias that could have been responsible for some portion of our results. We addressed this concern by mirror-flipping the hill, such that the slope of the hill ran from left to right.

Amazon Mechanical Turk users ($N = 483$) were recruited from the United States to participate in this study in exchange for 0.15 USD. The experimental setup was exactly the same as

in Experiment 1, with the exception that the scene (hill) and car were horizontally flipped to run from right to left (slope = 0.35). In the process of flipping the image of the car on the y -axis, the car’s origin changed from the upper left to the upper right corner. The position of the car was also adjusted to be in *roughly* the same position as in Experiment 1, relative to the current scene (410, -120). This adjustment was necessary to place the car on the road (an exact horizontal flip of the car placed it in the bushes of the scene).

Results and discussion

A 2 (Orientation: facing up/down) \times 3 (Language: “The car moved *forward*,” “The car moved *backward*,” or no language) between-subjects ANOVA was performed on placement times, confidence, and GD placements. Before analysis, the actual location of the car on the GD ($\mu = 436.3$) was subtracted from the GD placements, centering the data on zero.

The data from 32 (<7%) participants were excluded from all following analyses due to responses greater than three standard deviations away from the mean for placement times, confidence, and/or GD placements. No main effect of language or orientation was apparent on GD placements. However, there was a significant interaction, $F(1, 455) = 12.72$, $p < .001$, $\eta_p^2 = .05$. Simple-effects analysis with Tukey HSD correction revealed that congruent conditions were significantly different from incongruent conditions ($ps < .01$ across all conditions). No other conditions were significantly different from one another (see Fig. 4).

Similar to the findings from Experiment 1, the placements of the car in both conditions in which language and orientation were

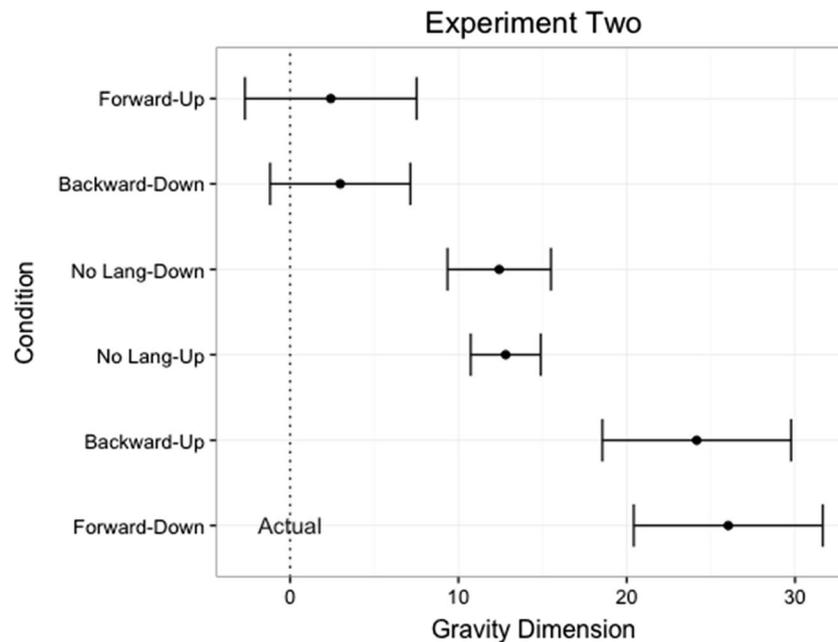


Fig. 4 Means and standard error bars for the remembered location of the car on the gravity dimension, by condition for Experiment 2. The dotted line indicates the actual position of the car, and higher values indicate stronger effects of implied gravitational motion on spatial memory

incongruent with gravity were not significantly different from the actual location of the car (“forward”–up: $M = 2.42$, $SE = 5.10$, $t(77) = -0.47$; “backward”–down: $M = 2.98$, $SE = 4.17$, $t(81) = .72$). Again replicating well-known effects of implied gravity in visual memory, and the results from Experiment 1, in both conditions in which no language was presented, cars were displaced significantly farther from the actual location in the direction of implied gravity [no-language–up: $M = 12.81$, $SE = 2.08$, $t(72) = 6.12$, $p < .001$; no-language–down: $M = 12.42$, $SE = 3.08$, $t(76) = 4.04$, $p < .001$]. Finally, spatial memory for the location of the car in conditions in which language and orientation were congruent with gravity was also displaced significantly beyond the actual location in the direction of implied gravity [“forward”–down, $M = 4.64$, $SE = 5.62$, $t(68) = 4.64$, $p < .001$; “backward”–up, $M = 24.16$, $SE = 5.61$, $t(71) = 4.31$, $p < .001$].

Placement times We found a main effect of language, $F(2, 445) = 24.41$, $p < .001$, $\eta_p^2 = .10$, on placement times, such that when language was present, participants were slower to place the car than when language was not present: no language ($M = 1,697.50$, $SE = 44.21$ ms) < “backward” ($M = 2,202.70$, $SE = 64.85$ ms; $p < .001$) and “forward” ($M = 2,223.57$, $SE = 68.86$ ms; $p < .001$). There were no other significant effects of placement times ($M = 20,414.59$, $SE = 365.54$). A linear regression analysis in R (lm) with placement times as a predictor revealed no main effect of placement times on GD: $t(449) = 0.47$, $p = .47$

Confidence We observed a main effect of orientation, $F(1, 445) = 4.40$, $p = .04$, $\eta_p^2 = .01$, such that when the car was facing up the hill ($M = 3.5$, $SE = 0.13$), participants were more

confident in their placements than when the car was facing down hill ($M = 3.12$, $SE = 0.12$). Neither language ($p = .49$) nor an interaction between language and orientation ($p = .113$) significantly predicted placement confidence ($M = 3.32$, $SE = 0.11$). Nor was confidence a significant predictor of GD placements: $t(454) = -1.78$, $p = .08$.

General discussion

Spatial memory appears to be, at least to a detectable extent, influenced by top-down information such as language. The direction of these effects appears to be “prospective”: that is, aligned in the implied or stated direction in which the object in the scene may move next. It is possible that motion language provides an independent contribution to spatial memory itself. Experiments 1 and 2 showed a robust interaction between language and image orientation, such that the directionality of implied motion in language was highly dependent on the visual presentation of the direction the car was facing. Additionally, the no-language conditions revealed that the orientation of the car had no effect when it was presented without language, suggesting that additional information was necessary to make certain visual information relevant to encoding location. Finally, motion language provides an implied motion effect on spatial memory beyond that of implied motion in vision. This suggests that prior information could influence encoding in ways that appear to enhance one’s ability to predict some future state.

These results are not inconsistent with a speculative perspective that vision integrates prior linguistic information

directly (Gennari et al., 2002; Lupyan, 2012; Lupyan et al., 2010; Meteyard et al., 2007; Spivey & Geng, 2001; Tanenhaus et al., 1995). Unfortunately, given the nature of the task, it was not possible to determine whether our effects were the result of language influencing online perceptual processes, such as visual perception, directly, or of postperceptual integration of distributed information. For example, it could be that verbal overshadowing—participants repeating language-related instructions—may occur at the point of recall (Alogna et al., 2014; Schooler & Engstler-Schooler, 1990). This could suggest that postencoding retrieval is influenced by language. This might change the locus of our effects, but may nevertheless be interesting. Moreover, additional visual motion information such as the observation of the car actually moving up or down the hill may alter the impact of linguistic information on spatial memory. Recent research shows that with additional contextual information when watching dynamic motion (a movie clip) individuals are more likely to experience greater overall comprehension and decreased cognitive load (Loschky, Larson, Magliano, & Smith, 2015). In our study, because the image we present is static, it affords an opportunity for language to disambiguate possible movement direction in which a dynamic scene might not. If so, language that implies the speed of motion such as “the car bolted down the hill” may impact the amount of spatial displacement.

Further investigation will be necessary to tease out these confounds or others. For example, a verbal interference task during recall might help distinguish overshadowing versus encoding theories of our effects. In this work, we aimed to make first steps toward determining whether language can influence visual spatial memory for objects already in one’s visual array. Future work should aim to replicate the present findings using various static images and scenes (such as changing the angle of the road) as well as modifying the scene such that the perspective of the viewer and the car is more directly aligned with one’s true visual experience (e.g., making sure the scene does not appear visually awkward as may be the case to some participants in the present task due to slight misalignment of road and car perspective).

Finally, another potential confound was that the task might simply have modulated participants’ strategies offline before viewing the image, similar to not perceiving certain aspects of a scene due to prior goals (Drew et al., 2013; Simons & Chabris, 1999). This alternative explanation seems less compelling to us, given how comprehension unfolds in the task itself. For an offline strategy to be feasible, it would be necessary to know the intended *direction* of motion prior to viewing the image of the car (e.g., up, down, left, and right; see Meteyard et al., 2007; Spivey & Geng, 2001). In the present study, the car’s movement implied by language—*forward* or *backward*—could be in any actual direction (left, right, up, or down) until further elucidated by additional (visual)

information, such as the car’s facing direction, present in the image itself. Furthermore, the use of only a single stimulus in our experiment helped control for possible expectations of motion that might have stemmed from the presentation of multiple stimuli. Again, only when implied motion is present in vision can participants fully comprehend the true direction of motion in language.

The time course of visual encoding would seem to afford the integration of top-down (linguistic) information. In the present task, global visual information such as the scene itself might first be experienced, followed by the reorganization (within milliseconds) of the same neural mechanisms used to process initial contextual information toward experiencing finer perceptual information (Sugase et al., 1999), such as the car’s orientation. The time between neuronal firing allows additional global information, such as prior linguistic information, to propagate through the cortex and recursively bias encoding and recall with relevant top-down information (Lamme & Roelfsema, 2000). Such integrative processes during encoding and retrieval would, presumably, help the visual system interpret finer details of a scene relative to its current context. It is possible that distributed—visual and linguistic—information activates the same neural regions to integrate motion information within the same perceptual experience. Put another way, the unfolding of visual perception over time affords an opportunity for nonspecific implied motion in language to be comprehended relative to—and within the same process as—visual motion perception. It is unlikely that implied motion present in language induces an offline strategy that can be used to modulate visual perception in this task.

Yet another possible explanation is that motion language may have directly influenced ocular-motor action (Spivey & Geng, 2001). Implied motion in vision and language may prime participants to orient their gaze toward the front or rear of the car, an effect that could influence the remembered location of the car itself (see Kerzel, 2000). However, if eye movements alone were to account for the observed effects, we would anticipate that no difference would occur between conditions in which the directionality of eye movements were the same. That is, in conditions in which implied motion was congruent with gravity and no-language conditions the eyes would be pulled in the same direction, toward the bottom of the hill. However, when motion language was present, the implied effects of motion on spatial memory were much greater than those implied by vision alone. In addition, for the participant to understand where to move the eyes when provided with language—either to the front or rear of the car—he or she would have to know which direction the car was facing. This information was not available until the car itself was presented. Therefore, to fully comprehend the direction of motion present in language, it was necessary that the participant first perceive the visual stimulus, at which time implied motion in vision was also present. Given that the full

effect of motion in language occurs only when implied motion in vision is present, it is more reasonable to conclude that participants' eye movements were influenced by implied motion from language and implied motion from vision *simultaneously*. Thus, the effects of motion language and vision appear to be distributed and additive: Regardless of the eyes being pulled, the effects on spatial memory involve influences from *both* vision and language, at the same time.

Our findings highlight the integrative process of linguistic comprehension and visual perception and provide further support for theories that conceptualize the purpose of cognition as being at least partly about predicting the outcomes of actions and events (Bar, 2009; Bubic et al., 2010; Clark, 2013a, b; Hommel et al., 2001; Jordan, 2013; Neisser, 1976; Schubotz, 2007; Spratling, 2010). Our findings are commensurate with previous research that suggested that visual processing is influenced by top-down information (Büchel & Friston, 1997; Friston, 2005; Goldstone et al., 2015; Jordan & Hunsinger, 2008; Lupyán, 2012; Lupyán & Clark, 2015; Stins & van Leeuwen, 1993; Ullman, 1995) and that simulations in language and vision are unified in what might be a dynamic situation model (Zwaan & Radvansky, 1998). We found evidence that prior linguistic information does not simply mediate the implied-motion information present in vision, controlling what one might experience in visual perception, but that it provides additional information that is integrated into the perceptual system. It is likely that *more* information, regardless of how it is realized (e.g., through vision, linguistic comprehension, or current or past actions) enhances our ability to predict the outcomes of current events.

Conclusion

In two experiments, we aimed to test whether linguistic information influences the perceptual process in adaptive ways. Our findings support a notion of perception that involves the integration of information from vision and language, with the adaptive advantage of enhancing our ability to predict event outcomes. Typically thought of as errors in visual memory, misremembering the actual location of the objects of a scene might instead be better *reconceptualized* as the prediction of future events, an effect that can be measured through spatial memory. Provided that time does not stop, the *outcomes* of events are necessarily in the future. At any given moment, how we benefit from more information may be to get as far ahead as possible.

Author note We thank three anonymous reviewers and the action editor for their helpful comments and critiques. This work was supported in part by an IBM PhD Fellowship (2015/16) awarded to D.W.V.

References

- Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Bornstein, B.,...Carlson, C. (2014). Contribution to Alonga et al. (2014). Registered replication report: Schooler & Engstler-Schooler (1990). *Perspectives on Psychological Science*, 9, 556–578.
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, 15, 600–609.
- Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B*, 364, 1235–1243.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M.,...Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103, 449–454. doi:10.1073/pnas.0507062103
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660. doi:10.1017/s0140525x99002149
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B*, 364, 1281–1289.
- Barto, A. G., Sutton, R. S., & Watkins, C. J. C. H. (1990). Learning and sequential decision making. In M. R. Gabriel & J. Moore (Eds.), *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 539–602). Cambridge, MA: MIT Press.
- Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4(25), 1–15. doi:10.3389/fnhum.2010.00025
- Büchel, C., & Friston, K. J. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex*, 7, 768–778.
- Clark, A. (2007). What reaching teaches: Consciousness, control, and the inner zombie. *British Journal for the Philosophy of Science*, 58, 563–594. doi:10.1093/bjps/axm030
- Clark, A. (2013a). Expecting the world: Perception, prediction, and the origins of human knowledge. *Journal of Philosophy*, 110, 469–496.
- Clark, A. (2013b). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181–204. doi:10.1017/S0140525X12000477
- Coventry, K. R., Christophel, T. B., Fehr, T., Valdés-Conroy, B., & Herrmann, M. (2013). Multiple routes to mental animation: Language and functional relations drive motion processing for static images. *Psychological Science*, 24, 1379–1388. doi:10.1177/0956797612469209
- De Sá Teixeira, N. A., Hecht, H., & Oliveira, A. M. (2013). The representational dynamics of remembered projectile locations. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 1690–1699. doi:10.1037/a0031777
- Drew, T., Vö, M. L.-H., & Wolfe, J. M. (2013). The invisible gorilla strikes again sustained inattention blindness in expert observers. *Psychological Science*, 24, 1848–1853.
- Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27, 143–165. doi:10.1016/0749-596X(88)90071-X
- Flanagan, J. R., Vetter, P., Johansson, R. S., & Wolpert, D. M. (2003). Prediction precedes control in motor learning. *Current Biology*, 13, 146–150.
- Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics*, 33, 575–581.
- Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94, 427–438. doi:10.1037/0033-295X.94.4.427
- Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 126–132. doi:10.1037/0278-7393.10.1.126

- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, *360*, 815–836.
- Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*, 49–79.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. New York, NY: Psychology Press.
- Glenberg, A. M., Lopez-Mobilia, G., McBeath, M., Toma, M., Sato, M., & Cattaneo, L. (2010). Knowing beans: Human mirror mechanisms revealed through motor adaptation. *Frontiers in Human Neuroscience*, *4*, 206. doi:10.3389/fnhum.2010.00206
- Goldstone, R. L., de Leeuw, J. R., & Landy, D. H. (2015). Fitting perception in and to cognition. *Cognition*, *135*, 24–29.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 849–878, disc. 878–937. doi:10.1017/S0140525X01000103
- Hubbard, T. L. (1995). Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic Bulletin & Review*, *2*, 322–338. doi:10.3758/BF03210971
- Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, *12*, 822–851. doi:10.3758/BF03196775
- Hubbard, T. L. (2014). Forms of momentum across space: Representational, operational, and attentional. *Psychonomic Bulletin & Review*, *21*, 1371–1403. doi:10.3758/s13423-014-0624-3
- Hubbard, T. L., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, *44*, 211–221. doi:10.3758/BF03206290
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic–connectionist theory of relational inference and generalization. *Psychological Review*, *110*, 220–264. doi:10.1037/0033-295X.110.2.220
- Jones, M., Curran, T., Mozer, M. C., & Wilder, M. H. (2013). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review*, *120*, 628–666. doi:10.1037/a0033180
- Jordan, J. S. (2013). The wild ways of conscious will: What we do, how we do it, and why it has meaning. *Frontiers in Psychology*, *4*, 574. doi:10.3389/fpsyg.2013.00574
- Jordan, J. S., & Hunsinger, M. (2008). Learned patterns of action-effect anticipation contribute to the spatial displacement of continuously moving stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 113–124. doi:10.1037/0096-1523.34.1.113
- Kerzel, D. (2000). Eye movements and visible persistence explain the mislocalization of the final position of a moving target. *Vision Research*, *40*, 3703–3715.
- Kinsbourne, M., & Jordan, J. S. (2009). Embodied anticipation: A neurodevelopmental interpretation. *Discourse Processing*, *46*, 103–126. doi:10.1080/01638530902728942
- Körding, K. P., & König, P. (2000). Learning with two sites of synaptic integration. *Network: Computation in Neural Systems*, *11*, 25–39.
- Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, *65*, 145–168.
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*, 571–579. doi:10.1016/S0166-2236(00)01657-X
- Lieberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current Biology*, *24*, 2569–2574. doi:10.1016/j.cub.2014.09.025
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 585–589.
- Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would Jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLoS ONE*, *10*, e0142474. doi:10.1371/journal.pone.0142474
- Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54.
- Lupyan, G., & Clark, A. (2015). Words and the world predictive coding and the language–perception–cognition interface. *Current Directions in Psychological Science*, *24*, 279–284.
- Lupyan, G., Thompson-Schill, S. L., & Swingle, D. (2010). Conceptual penetration of visual processing. *Psychological Science*, *21*, 682.
- Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition*, *31*, 663–672. doi:10.3758/BF03196106
- Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition*, *32*, 1389–1400. doi:10.3758/BF03206329
- Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs language affects low-level visual perception. *Psychological Science*, *18*, 1007–1013.
- Morrow, D. G., & Clark, H. H. (1988). Interpreting words in spatial descriptions. *Language and Cognitive Processing*, *3*, 275–291.
- Morsella, E., Godwin, C. A., Jantz, T. K., Krieger, S. C., & Gazzaley, A. (2015). Homing in on consciousness in the nervous system: An action-based synthesis. *Behavioral and Brain Sciences* Advance online publication. doi:10.1017/S0140525X15000643
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. New York, NY: W. H. Freeman.
- Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science*, *1*, 110–122. doi:10.1111/j.1745-6916.2006.00008.x
- Qian, T., & Aslin, R. N. (2014). Learning bundles of stimuli renders stimulus order as a cue, not a confound. *Proceedings of the National Academy of Sciences*, *111*, 14400–14405.
- Radvansky, G. A., Zwaan, R., Federico, T., & Franklin, N. (1998). Retrieval from temporally organized situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1224–1237. doi:10.1037/0278-7393.24.6.1224
- Reed, C. L., & Vinson, N. G. (1996). Conceptual effects on representational momentum. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 839–850.
- Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition*, *3*, 83–119.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*, 36–71. doi:10.1016/0010-0285(90)90003-M
- Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences*, *11*, 211–218.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, *28*, 1059–1074. doi:10.1068/p2952
- Skipper, J. I. (2014). Echoes of the spoken past: How auditory cortex hears context during speech perception. *Philosophical Transactions of the Royal Society B*, *369*, 20130297.
- Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*, 235–241. doi:10.1007/s004260100059
- Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *Journal of Neuroscience*, *30*, 3531–3543.
- Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*, 153–156. doi:10.1111/1467-9280.00326
- Stins, J. F., & van Leeuwen, C. (1993). Context influence on the perception of figures as conditional upon perceptual organization strategies. *Perception & Psychophysics*, *53*, 34–42.

- Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869–873.
- 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. doi:10.1126/science.7777863
- Ullman, S. (1995). Sequence seeking and counter streams: A computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, *5*, 1–11.
- Vinson, D. W., Abney, D. H., Dale, R., & Matlock, T. (2014). High-level context effects on spatial displacement: The effects of body orientation and language on memory. *Frontiers in Psychology*, *5*, 637. doi:10.3389/fpsyg.2014.00637
- Vinson, D. W., Abney, D. H., Amso, D., Chemero, A., Cutting, J. E., Dale, R.,...Spivey, M. J. (2016). Perception, as you make it. *Behavioral and Brain Sciences*. Advance online publication.
- Vinson, D. W., Jordan, J. S., & Hund, A. M. (2017). Perceptually walking in another's shoes: Goals and memories constrain spatial perception. *Psychological Research*, *81*, 66–74. doi:10.1007/s00426-015-0714-5
- Witt, J. K. (2011). Action's effect on perception. *Current Directions in Psychological Science*, *20*, 201–206. doi:10.1177/0963721411408770
- Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35–62). San Diego, CA: Academic Press.
- Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS ONE*, *7*, e51382. doi:10.1371/journal.pone.0051382
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185. doi:10.1037/0033-2909.123.2.162
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*, 168–171.