

Implied Motion Language Can Influence Visual Spatial Memory

David. W. Vinson^a, Jan Engelen^b, Rolf A. Zwaan^b, Teenie Matlock^a & Rick Dale^a

^aUniversity of California, Merced, Merced, CA

^bErasmus University of Rotterdam, Netherlands

Address Correspondence to:

David W. Vinson

dvinson@ucmerced.edu

University of California, Merced

5200 N. Lake Rd. Merced, CA. 95343

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 three experiments we tested whether the prior presentation of motion language influences visual spatial memory in ways that afford greater perceptual prediction. Experiment One found that motion language influenced judgments for the spatial memory of an object that implied motion. Experiment Two replicated this finding. A summary analysis comparing Experiment One and Experiment Two shows how motion language influences judgments of spatial memory beyond the implied motion present in the image itself. Our findings support a theory of perception as prediction.

Keywords: Implied motion, Motion language, Spatial cognition, Visual memory, Prediction

Word count: 5734

Implied Motion Language Can Influence Visual Spatial Memory

Imagine that you are driving down a highway, and seeing 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, Johansson, & Wolpert, 2003). The strength of the relationship between motor action 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, 1989; Bubic, Von Cramon & Schubotz, 2010; Clark 2007; Clark, 2013a; Clark, 2013b; 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 review), guided by past experiences (Bar, 2009; Barsalou, 1999, 2009), in evolutionarily adaptive ways (Gibson, 1979; Proffitt, 2006; Witt, 2011). In this view, the constant stream of visual information is influenced by a variety of prior *top-down* information (Goldstone, Leeuw, & Landy, 2015; Kveraga, Ghuman, & Bar, 2007; Lamme & Roelfsema, 2000; Lupyan & Clark, 2015; Stins & Van Leeuwen, 1993; Sugase, Yamane, Ueno, & Kawano, 1999) including conceptual and categorical labeling (Lupyan, 2012; Lupyan, Thompson-Schill, & Swingley, 2010),

language (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Skipper, 2014), goals (Büchel & Friston, 1997) and past actions (e.g., memories, Friston, 2005; Hummel & Holyoak, 2003; Jones, Mozer, & Wilder, 2013; Jordan & Hunsinger, 2008).

If prediction is ubiquitous in everyday cognition, then it should operate in many domains, not just simple perception. Our *memory* for a scene or event should also be influenced by the dynamics of prediction. In the current study, we examine the influence of top-down information on visual spatial 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, 2015). 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 trying to count the number of passes made by basketball players (Simons & Chabris, 1999; see also Drew, Vo, &

Wolfe, 2013). Prior stimuli can also impact next judgments. Participants tend to remember features of faces as being more similar to those previously seen (Lieberman, 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, 2013b). 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 current case, we expect that spatial memory for an object can be displaced predictably by how it is implied to be moving in the scene. In our task, the car does *not* move in the static scene, but language in the instructions may 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 (Madden & Zwaan, 2003; Glenberg et al., 2010; Santos, 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 perceptual threshold when 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 (for instance, 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 compared to 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, 1985), implied gravity (Hubbard & Bharucha, 1988; De Sá Teixeira, Hecht, & Oliveira, 2013) and stimulus orientation (Vinson, Abney, Matlock, & Dale, 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 detectible via post-perceptual 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 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. Further, 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 the effects of implied motion present in vision alone. If so, when language is congruent with visual implied motion, the remembered location of the object will be displaced farther along its implied motion trajectory than when only presented with visual implied motion. Moreover, when 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 effectively cancel out the effects of one another.

Current 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 current method involves the presentation of a stimulus only once. This is novel in that all other experimental designs in implied motion consist of multiple sequentially presented stimuli, that when presented together imply motion (see Hubbard 2005; 2014 for review).¹ In these experiments the location of the stimulus, when acted upon solely by implied gravity, is remembered as farther along an implied gravitational trajectory. However, implied vertical motion is also influenced by conceptual knowledge about the stimulus (Vinson & Reed, 1996). If participants anticipate the car's movement to be in the direction the car is facing, then 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 even farther still when language congruent with implied gravity is presented. When language is incongruent the effects of visual implied gravity may not occur.

We first present 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. 1). If there is an effect of visually implied gravitational motion

¹ Freyd (1983) presented participants with a single image of an object followed by a test image. However, the test image presented 250ms later contained that object farther along

alone, then the car should be remembered as farther down the hill in the No Language condition. This condition acts as a simple control replicating previous, well-known implied gravitational motion effects. Both experiments tested whether motion language influences 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, when participants read a sentence that is incongruent with visual implied motion, such as “The car moved *forward*” when the car is facing *Up* hill or “The car moved *backward*” when the car is facing *Down* hill, there will be no effect of implied motion on spatial memory. In this sense, motion language would effectively cancel out any effects of implied motion in vision. This may be the result of the presence of incongruent simulations occurring simultaneously or the lack of a sufficient congruent simulation due to incongruent information from vision and language. If language only mediates the effects of vision of spatial memory, when motion language is congruent with implied visual motion, such as “The car moved *forward*” when the car is facing *Down* hill or “The car moved *backward*” when the car is facing *Up* hill 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 displaced farther than the effects of implied motion from vision alone. If so, the effect of implied motion in congruent conditions will be greater than implied motion present in vision alone (i.e., in the No Language conditions).

Method

A total of 479 Amazon Mechanical Turk² users from the United States participated in exchange for \$0.15 *USD*. Roughly 75 individuals participated in each of six conditions, replicating 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: davevinson.com/exp/vzm/2/z2.html.

Using Adobe Flash CS6, participants observed 8 scenes, each with content that differed from the previous scene (See Fig. 1). Scenes were presented relative to the Adobe Flash stage *width X height* dimensions (550 X 400 pixels). The origin of the scene was located at the top-left corner of the screen, thus x-coordinates are positive and y-coordinates negative. The upper-left corner of each object was used as the object's origin to the x-, y-coordinate. The image of the car described below indicates the position of its upper-left corner on the scene unless otherwise specified.

² Amazon Mechanical Turk is a crowdsourcing marketplace where individuals are compensated for their participation in tasks setup by researchers and companies.

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 2000ms containing a white background with one of the two sentences: "*The car moved forward*" or "*The car moved backward*". A single text box (550 x 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 1000ms duration. Scene 4 presented for 1000ms contained the test image of a San Francisco street hill (slope = -0.35) and a red car (275 x 85) on the street with its origin (upper left corner) located at (120,-120). Scene 5 contained a black backdrop mask identical to Scene 3 both in presentation and duration. Both black backdrop masks replicated previous experimental masking times used in recent spatial displacement studies (De Sá Teixeira, Hecht, & Oliveira, 2013).

Scene 6 was identical to Scene 4 except the car was omitted from the scene. The cursor, having disappeared at the presentation of Scene 2, re-appeared but this time, instead of a mouse, it was the exact image of the test stimulus from Scene 4 (e.g., the car), oriented the same way it was in Scene 4. The precise position of the participant's x-,y coordinate was the upper left corner, the origin, of the car. Because participants did not see their mouse, no other point of reference was present that could have alluded to where exactly their cursor was located. Participants were to place the car by clicking on the screen where they remembered seeing it last, as indicated by previous instructions. The initial position of the car (i.e. participant's mouse) was not controlled; wherever their mouse was 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 current authors has 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) X 3 (Language: “The car moved *forward*” / “The car moved *backward*” / No Language) between subjects ANOVA was performed on placement times, confidence and a gravity dimension (G-dimension) determined by transforming the *xy*-coordinate plane onto a single dimension aligned to the slope of the hill (see Fig. 2). The transformation results in a new coordinate for the actual location of the car on the gravity dimension ($\mu = 127.7$) and allows for a more direct assessment of the impact of implied gravitational motion on participant placements. Responses more than three standard deviations away from the mean of placement times, confidence and/or the G-dimension were removed from all subsequent analyses (a total of 25 participants, ~5%).

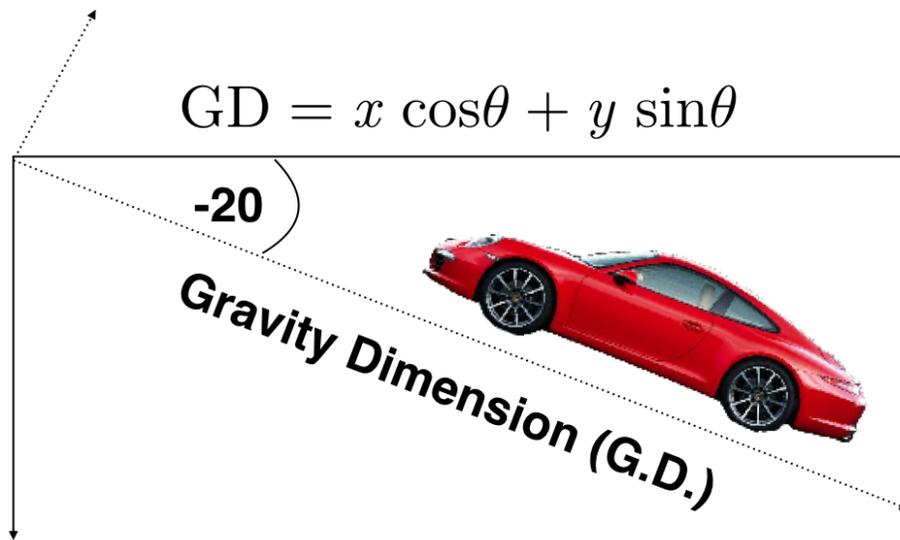


Fig 2. Coordinate transformation from the original xy -coordinate plane to the G-dimension. This specific example reflects the G-dimension and angle from Experiment One.

Gravity Dimension. Before analysis, the actual location of the car on the G-dimension was subtracted from participant placements centering the data on zero. There was no main effect of Language or Orientation on G-dimension placements, but there was a significant interaction $F(1,448) = 27.13, p < .001, \eta_p^2 = .11$ (Fig. 3). Simple effects analysis with TukeyHSD correction revealed conditions where Language and Orientation were congruent with the direction of gravity were significantly farther displaced in the direction of gravity than both no language conditions and conditions where Language and Orientation were incongruent with gravity while the no language conditions were also significantly different than conditions that were incongruent with gravity ($p < .001$; with

the exception of the comparison between “*Backward*”-Up and no language-Down, $p = .08$).

Spatial memory for the location of the car in conditions where Language and Orientation were congruent with gravity was significantly displaced beyond the actual location on the gravity dimension. Both conditions where Language and Orientation were incongruent with gravity were not significantly different from the actual location of the car (“Forward”-Up: $M = -1.53$, $SE = 5.00$, $t(70) = -0.31$; “Backward”-Down: $M = -5.20$, $SE = 3.79$, $t(80) = -1.37$). Both conditions where no language was presented were significantly farther displaced from the actual location of the car in the direction of implied gravity (No Language-Up: $M = 9.00$, $SE = 4.22$, $t(76) = 2.13$, $p = .04$; No Language-Down: $M = 10.19$, $SE = 3.99$, $t(75) = 2.56$, $p = .01$) replicating well-known effects of implied gravity in visual memory. Finally, both conditions where Language and Orientation were congruent with gravity were significantly farther displaced in the direction of gravity from the actual location (“Backward”-Up: $M = 37.7$, $SE = 6.64$, $t(77) = 5.68$, $p < .001$; “Forward”-Down: $M = 29.61$, $SE = 5.88$, $t(70) = 5.03$, $p < .001$).

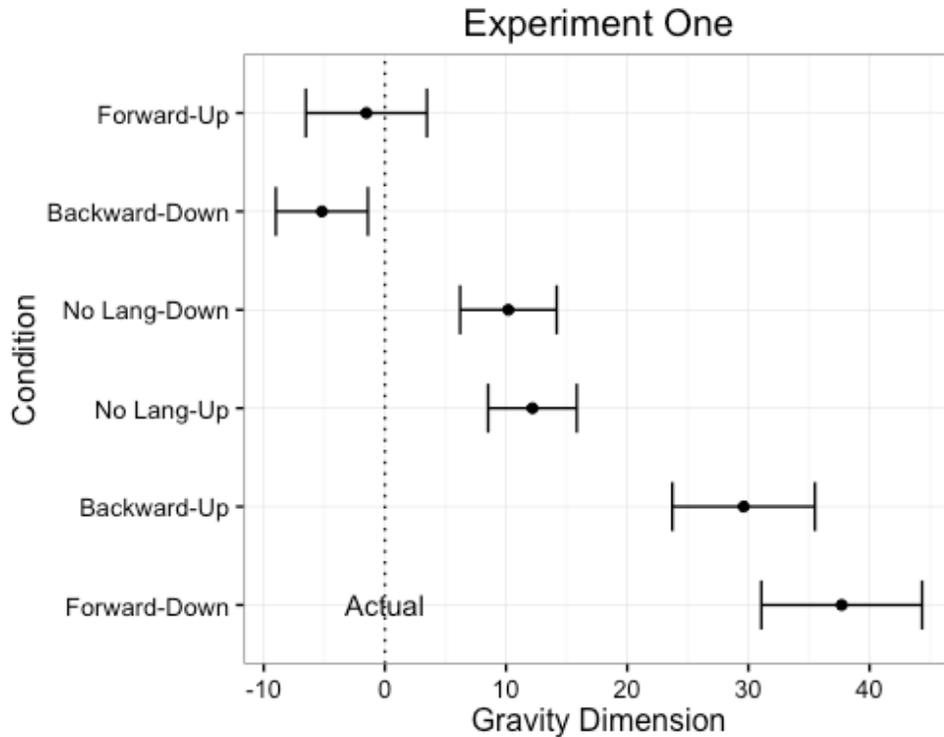


Fig 3. The mean and standard error bars of the remembered location of the car on the G-dimension by Condition for Experiment One. The dotted line indicates the actual position of the car. Higher values indicate stronger effects of implied gravitational motion on spatial memory.

Placement times. The experimental paradigm was not designed to test placement times. They are reported here solely for the sake of completeness. There was a main effect of Language, $F(2,448) = 5.18, p = .006$ for placement times such that when language was present, participants were slower to place the car than when language was not present: No Language ($M = 1912.58, SE = 55.53\text{ms}$) < “Backward” ($M = 2105.70, SE = 52.32\text{ms}; p = .04$) and “Forward” ($M = 2148.83, SE = 58.23\text{ms}; p = .008$). While previous research shows the effects of implied gravitational motion on stimulus

placement increases with longer delays between image observation and participant responses (De Sá Teixeira, Hecht & Oliveira, 2013), we did not predict that longer placement times would result in greater displacement as all participants experienced the same inter-stimulus-interval (1000ms). A linear regression analysis in R (lm) with placement times as a predictor of G-dimension placements revealed no main effect of placement times on the G-dimension; $F(1,452) = .06, p > .05$

Confidence. Neither Language nor Orientation or an interaction between Language and Orientation significantly predicted placement confidence ($M = 3.22, SE = .09$). Nor was confidence a significant predictor of G-dimension placements; $F(1,452) < .001, p > .05$.

Experiment Two

In Experiment One, 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, which is responsible for some portion of our results. We addressed this concern by mirror-flipping the hill, such that the slope of the hill runs 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 Experiment One with the exception that the scene (hill) and car were horizontally flipped to run from right to left (slope = .35). In the process of flipping the image of the car on the y-axis, the car's origin changed from the upper left corner, to upper right. The position of the car was also adjusted to be in *roughly* the same position

as Experiment One 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 the car in the bushes of the scene).

Results and Discussion. A 2 (Orientation: facing Up / Down) X 3 (Language: “The car moved *forward*”, “The car moved *backward*” or No Language) between subjects ANOVA was performed on placement times, confidence and G-dimension placements. Before analysis, the actual location of the car on the G-dimension ($\mu = 426.2$) was subtracted from G-dimension placements centering the data on zero.

Data from 23 (< 5%) participants were excluded from all following analyses due to responses greater than three standard deviations away from the mean of placement times, confidence and/or G-dimension placements. There was no main effect of Language or Orientation on G-dimension placement. However, there was a significant interaction $F(1,450) = 15.88, p < .001, \eta_p^2 = .07$. Simple effects analysis with TukeyHSD correction revealed that congruent conditions were significantly different from incongruent conditions ($p < .001$ across all conditions). No other conditions were significantly different from one another (see Fig. 4).

Similar to findings from Experiment One, both conditions where Language and Orientation were incongruent with gravity were not significantly different from the actual location of the car (“Forward”-Up: $M = .21, SE = 5.69, t(79) = -0.04$; “Backward”-Down: $M = 4.63, SE = 4.51, t(82) = 1.03$). And again replicating well-known effects of implied gravity in visual memory and results from Experiment One, both conditions where no language was presented were significantly farther displaced from the actual location of

the car in the direction of implied gravity (No Language-Up: $M = 13.41$, $SE = 2.02$, $t(72) = 6.65$, $p < .001$; No Language-Down: $M = 13.58$, $SE = 3.34$, $t(76) = 4.07$, $p < .001$). Finally, spatial memory for the location of the car in conditions where Language and Orientation were congruent with gravity was significantly displaced beyond the actual location in the direction of implied gravity (“Forward”-Down, $M = 4.87$, $SE = 5.76$, $t(68) = 4.88$, $p < .001$; “Backward”-Up, $M = 30.43$, $SE = 6.14$, $t(73) = 4.95$, $p < .001$).

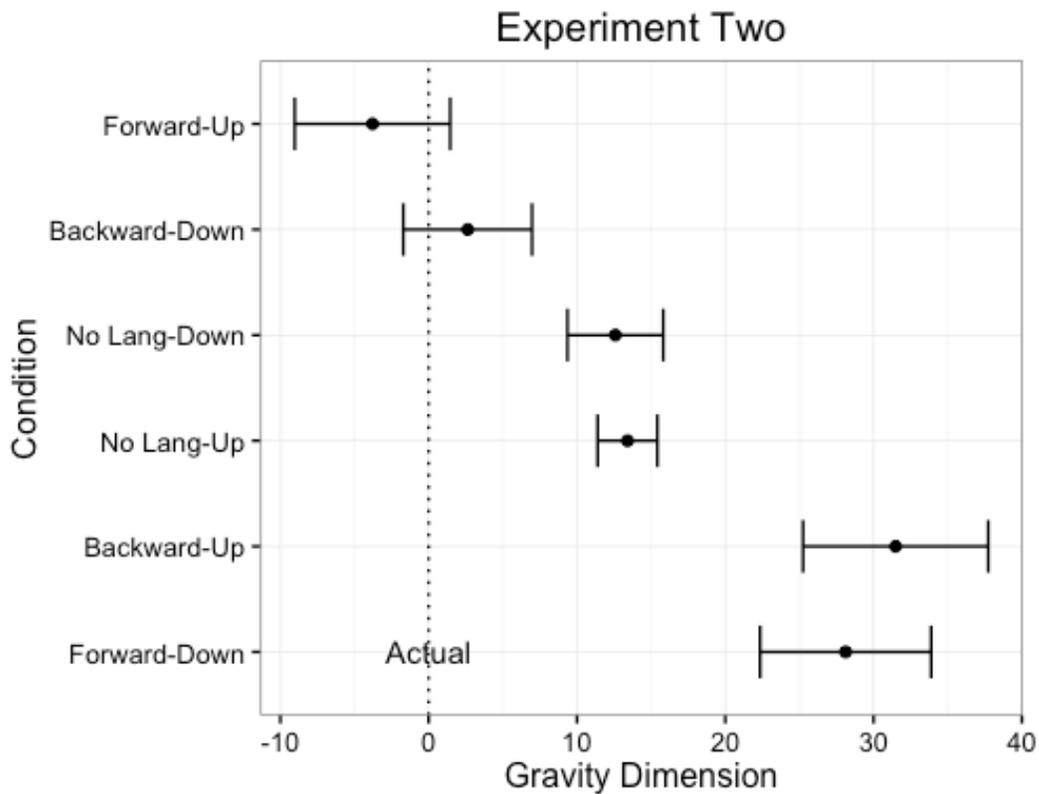


Fig 4. The mean and standard error bars of the remembered location of the car on the G-dimension by Condition for Experiment Two. The dotted line indicates the actual position of the car. Higher values indicate stronger effects of implied gravitational motion on spatial memory.

Placement times. No main effects or interactions for placement times ($M = 2043.39$, $SE = 36.32\text{ms}$) for Language or Orientation were observed. A linear regression analysis in R (lm) with placement times as a predictor revealed no main effect of placement times on the G-dimension: $t(454) = -1.49$, $p = .14$

Confidence. There was a main effect of Orientation, $F(1,450) = 4.33$, $p = .04$, such that when the car was facing up the hill ($M = 3.5$, $SE = .13$) participants were more confident in their placement than when the car was facing down hill ($M = 3.13$, $SE = .12$). Neither Language ($p = .49$) nor an interaction between Language and Orientation ($p = .113$) significantly predicted placement confidence ($M = 3.32$, $SE = .11$). Nor was confidence a significant predictor of G-dimension placements; $t(454) = -1.78$, $p = .08$.

General Discussion

Spatial memory appears to be, at least to a detectable extent, labile, influenced by top-down information such as language. The direction of these effects appears to be “prospective”: 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 show a robust interaction between language and image orientation such that the directionality of implied motion in language is highly dependent on the visual presentation of the direction the car is facing. Additionally, No Language conditions found that the orientation of the car had no effect when 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.

Results are not inconsistent with a speculative perspective that vision is integrating prior linguistic information directly (Gennari, Sloman, Malt, & Fitch, 2002; Lupyan, 2012; Lupyan, Thompson-Schill & Swingley, 2010; Meteyard, Bahrami, & Vigliocco, 2007; Spivey & Geng, 2001; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995). Unfortunately, given the nature of the task, it is not possible to determine if our effects are the result of language influencing online perceptual processes, such as visual perception directly, or the post-perceptual 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 post-encoding 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 where a dynamic scene might not. If so, language that implies the speed of motion such as “*the car bolted down the hill*” may still impact the amount of spatial displacement.

Further investigation is necessary to tease out these confounds or others. For example a verbal interference task during recall might help distinguish overshadowing vs. encoding theories of our effects. In this work we aimed to make first steps toward determining if language can influence visual spatial memory for objects already in one's visual array. Future work should aim to replicate the current 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 current task due to slight misalignment of road and car perspective).

Lastly, another potential confound is that the task may simply modulate participants' strategies offline before viewing the image, similar to not perceiving certain aspects of a scene due to prior goals (Drew, Vo, & Wolfe, 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 is 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, Bahrami, & Vigliocco, 2007; Spivey & Geng, 2001). In the current study, the car's movement implied by language—*forward* or *backward*—could be 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. Further, the use of only a single stimulus in our experiment helps control for possible expectations of motion that stem from the presentation of multiple stimuli.

Again, it is only when implied motion in vision is present that participants can fully comprehend the true direction of motion in language.

In addition, the time course of visual encoding would seem to afford the integration of top-down (linguistic) information. In the current 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, Yamane, Ueno, & Kawano, 1999), such as the car's orientation. The time between neuronal firing allows 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 the results of this study may be the result of motion language directly influencing ocular-motor action (Spivey & Geng, 2001). Implied motion in vision and language may have primed participants to orient their gaze toward the front or rear of the car, an effect that could have influenced 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 there to be no difference between conditions where the directionality of eye movements were the same. In conditions where 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 implied motion in vision alone. In addition, for the participant to understand where to move their eyes when provided with language—either to the front or rear of the car—she would have to know which direction the car was facing. This information is not available until the car itself is presented. Therefore, to fully comprehend the *direction* of motion present in language it is necessary that the participant perceive the visual stimulus, at which time implied motion in vision is 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 are 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 to be at least partly about predicting the outcome of actions and events (Bar, 2009; Bubic, Von Cramon & Schubotz, 2010; Clark, 2013a, 2013b; Hommel, et al., 2001; Jordan, 2013; Neisser, 1976; Schubotz, 2007; Spratling, 2010). Our findings are

commensurate with previous research that suggests visual processing is influenced by top-down information (Büchel & Friston, 1997; Friston, 2005; Goldstone, Leeuw & Landy, 2015; Jordan & Hunsinger, 2008; Lupyan & Clark, 2015; Lupyan, 2012; 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 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, current or past actions) enhances our ability 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 objects of a scene might instead be better *re*-conceptualized as the prediction of future events; an effect that can be measured through spatial memory. Provided that time does not stop, *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.

Acknowledgements

We would like to 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 David W. Vinson.

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(5), 556-578.
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, 15(4), 600-609.
- Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 364, 1235–1243.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2), 449-454.
- 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 of London Series B: Biological Sciences*, 364, 1281-1289.
- Barsalou, L.W. (2010). Grounded cognition: past, present, and future. *Topics in Cognitive*

- Science*, 2, 716–724
- Barto, A. G., Sutton, R.S., & Watkins, C. J. C. H. (1990). Learning and sequential decision making. In M. Gabriel, J. Moore (Eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (pp. 539–602). MIT Press, Cambridge, MA.
- Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4, 1-15.
- Boulenger, V., Roy, A., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. (2006). Cross-talk between language processes and overt motor behavior in the first 200 msec of processing. *Journal of Cognitive Neuroscience*, 18(10), 1607-1615.
- 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(8), 768-778.
- Clark A. (2007). What reaching teaches: consciousness, control, and the inner zombie. *British Journal for the Philosophy of Science*, 58, 563–594
10.1093/bjps/axm030
- Clark, A. (2013a). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03), 181-204.
- Clark, A. (2013b). Expecting the world: perception, prediction, and the origins of human knowledge. *The Journal of Philosophy*, 110(9), 469-496.
- 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*,

0956797612469209. 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(6), 1690.
- Drew, T., Võ, M. L. H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: sustained inattention blindness in expert observers. *Psychological Science*, 24(9), 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(2), 143-165.
- Flanagan, J. R., Vetter, P., Johansson, R. S., & Wolpert, D. M. (2003). Prediction precedes control in motor learning. *Current Biology*, 13(2), 146-150.
- Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics*, 33(6), 575-581.
- Freyd, J. (1987). Dynamic mental representations. *Psychological Review*, 94, 427-438.
- Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 126.
- Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological Sciences*, 360(1456), 815-836.
- Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, 83(1), 49-79.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. New York, New NY: Psychology Press.

- Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain sciences*, 20(01), 41-50.
- 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.
- Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, 48(7), 905-922.
- Goldstone, R. L., de Leeuw, J. R., & Landy, D. H. (2015). Fitting perception in and to cognition. *Cognition*, 135, 24-29.
- Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, 12, 1–19.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). A theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24(05), 910-926.
- Hubbard, T. L., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, 44(3), 211-221.
- Hubbard T. (1995). Environmental invariants in the representation of motion: implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic Bulletin & Review*, 2, 322–338 10.3758/BF03210971
- Hubbard T. (2005). Representational momentum and related displacements in spatial memory: a review of the findings. *Psychonomic Bulletin & Review*, 12, 822–851 10.3758/BF03196775

- Hubbard, T. L. (2014). Forms of momentum across space: Representational, operational, and attentional. *Psychonomic Bulletin & Review*, *21*(6), 1371-1403.
- Huette, S., Winter, B., Matlock, T., & Spivey, M. (2012). Processing motion implied in language: Eye-movement differences during aspect comprehension. *Cognitive Processing*, *13*, 193-197.
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *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*(3), 628.
- 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*.
- 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*(1), 113.
- Kerzel, D. (2000). Eye movements and visible persistence explain the mislocalization of the final position of a moving target. *Vision Research*, *40*(27), 3703-3715.
- Kinsbourne M., & Jordan J. S. (2009). Embodied anticipation: a neurodevelopmental interpretation. *Discourse Processing*, *46*, 103–126 10.1080/01638530902728942
- Körding, K. P., & König, P. (2000). Learning with two sites of synaptic integration. *Network: Computation in Neural systems*, *11*(1), 25-39.
- Kourtzi, Z. (2004). But still, it moves. *Trends in Cognitive Sciences*, *8*, 47–49.
- Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/ MST by static images

- with implied movement. *Journal of Cognitive Neuroscience*, 12, 48–55.
- Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, 65(2), 145-168.
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11), 571-579.
- 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(5), 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(11), e0142474.
- Lieberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current Biology*, 24(21), 2569-2574.
- 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(4), 279-284.
- Lupyan, G., Thompson-Schill, S. L., & Swingle, D. (2010). Conceptual penetration of visual processing. *Psychological Science*, 21(5), 682.
- Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations?. *Memory & Cognition*, 31(5), 663-672.
- Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition*, 32(8),

1389-1400.

- Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs language affects low-level visual perception. *Psychological Science, 18*(11), 1007-1013.
- Morrow, D.G. & Clark, H.H. (1988). Interpreting words in spatial descriptions. *Language and Cognitive Processing, 3*(4), 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 [Target Article]*.
- Neisser, U. (1976). Cognition and reality: principles and implications of cognitive psychology. WH Freeman/Times Books/Henry Holt & Co.
- Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science, 1*(2), 110-122.
- Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience, 17*, 1–9.
- 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*(40), 14400-14405.
- Radvansky, G. A., Zwaan, R., Federico, T., & Franklin, N. (1998). Retrieval from temporally organized situation models. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 1224-1237.
- 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.
- Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Sciences*, 11(5), 211-218.
- Schubotz, R. I., & von Cramon, D. Y. (2001). Functional organization of the lateral premotor cortex: fMRI reveals different regions activated by anticipation of object properties, location and speed. *Cognitive Brain Research*, 11(1), 97-112.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception-London*, 28(9), 1059-1074.
- Skipper JI. 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(4), 235-241.
- Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *The Journal of Neuroscience*, 30(9), 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*(1), 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*(6747), 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*(5217), 1632-1634.
- Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, *5*(1), 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*.
- Vinson, D. W., Jordan, J. S., & Hund, A. M. (2015). Perceptually Walking in Another's Shoes: Goals and Memories Constrain Spatial Perception. *Psychological Research*, *79*(5), 1-9
- Witt, J. K. (2011). Action's effect on perception. *Current Directions in Psychological Science*, *20*(3), 201-206.
- Zwaan, R.A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B.H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 44, pp. 35–62). San Diego, CA: Academic Press.

- Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*, *18*(5), 229-234.
- Zwaan, R.A., Stanfield, R.A., & Yaxley, R.H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*, 168–171.
- Zwaan, R. A., & Pecher, D. (2012) Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS ONE* *7*(12): e51382.
doi:10.1371/journal.pone.0051382
- Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. *Grounding cognition: The role of perception and action in memory, language, and thinking*, 224-245.
- Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science*, *28*(4), 611-619.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162.