

Cognitive Science 31 (2007) 1–31  
Copyright © 2007 Cognitive Science Society, Inc. All rights reserved.

# Spatial and Linguistic Aspects of Visual Imagery in Sentence Comprehension

Ben Bergen<sup>a</sup>, Shane Lindsay<sup>b</sup>, Teenie Matlock<sup>c</sup>, Srini Narayanan<sup>d</sup>

<sup>a</sup>*Department of Linguistics, University of Hawaii at Manoa*

<sup>b</sup>*University of Sussex*

<sup>c</sup>*University of California, Merced*

<sup>d</sup>*International Computer Science Institute*

Q1

Q2

Received 25 May 2005; received in revised form 30 November 2006; accepted 30 November 2006

---

## Abstract

There is mounting evidence that language comprehension involves the activation of mental imagery of the content of utterances (Barsalou, 1999; Bergen, Chang, & Narayan, 2004; Bergen, Narayan, & Feldman, 2003; Narayan, Bergen, & Weinberg, 2004; Richardson, Spivey, McRae, & Barsalou, 2003; Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002). This imagery can have motor or perceptual content. Three main questions about the process remain under-explored, however. First, are lexical associations with perception or motion sufficient to yield mental simulation, or is the integration of lexical semantics into larger structures, like sentences, necessary? Second, what linguistic elements (e.g., verbs, nouns, etc.) trigger mental simulations? Third, how detailed are the visual simulations performed? A series of behavioral experiments address these questions, using a visual object categorization task to investigate whether up- or down-related language selectively interferes with visual processing in the same part of the visual field (following Richardson et al., 2003). The results demonstrate that either subject nouns or main verbs can trigger visual imagery, but only when used in literal sentences about real space—metaphorical language does not yield significant effects—which implies that it is the comprehension of the sentence as a whole and not simply lexical associations that yields imagery effects. This article also finds that the evoked imagery contains detail as to the part of the visual field where the described scene would take place.

*Keywords:* Linguistics; Psychology; Communication; Language understanding; Mental simulation; Perception; Semantics; Human Experimentation; Spatial cognition; Abstract concepts

## 1. Introduction

“Thought is impossible without an image.” (Aristotle, *On Memory and Recollection*)

Q3

---

Correspondence should be addressed to Ben Bergen, Department of Linguistics, University of Hawaii at Manoa, 569 Moore Hall, 1890 East-West Rd., Honolulu, HI 96822. E-mail: bergen@hawaii.edu

Until the late 1950s, mental imagery was believed to occupy a special place in human thought. Throughout most of the second half of the 20th century, however, imagery was backgrounded by approaches that favored objectivism and symbol manipulation. Over the course of the past 2 decades, imagery has once again become increasingly more interesting to cognitive scientists. A number of studies have shown that humans automatically and unconsciously engage perceptual and motor imagery when performing high-level cognitive tasks, such as recall (Nyberg et al., 2001) and categorization (Barsalou, 1999). The benefit of conscripting imagery for these tasks is clear—imagery provides a modality-specific, continuous representation well suited for comparing with perceptual input or performing inference. Three scholarly traditions have converged on the notion that language understanding critically engages the cognitive capacity to internally construct modal representations. Cognitive linguistics, for one, has long emphasized the importance of embodied representations of the world (e.g., spatial topology) in the representation of language (e.g., Lakoff, 1987; Langacker, 1987). Embodied cognitive psychology has similarly highlighted the importance of low-level perceptual and motor processes in language and other high-level phenomena (Barsalou, 1999; Glenberg & Robertson, 2000). And research on mental models in narrative comprehension has emphasized the role of detailed perceptual and motor knowledge in the construction of mental representations of scenes from verbal input (Zwaan, 1999). This convergence of views has spawned several lines of empirical and theoretical work arguing that understanding language leads to the automatic and unconscious activation of mental imagery corresponding to the content of the utterance. Such imagery, which may be motor or perceptual in nature (among others), has the potential to interfere with (Kaschak et al., 2005; Richardson et al., 2003) or facilitate (Glenberg & Kaschak, 2002; Zwaan et al., 2002) the actual performance of actions or the perception of objects, depending on the task.

Q4

Q5

This article focuses on the role of visual imagery in language understanding, and provides evidence that language processing drives location-specific perceptual images of described entities and their attributes. It advances the study of language-induced mental simulation in three ways. First, previous work on mental imagery and language understanding has not explored which linguistic elements—nouns, verbs, or others—engage imagery in the course of understanding a sentence. The work reported here demonstrates that mental imagery can be evoked by either subject nouns or main verbs in sentence stimuli. Second, the work reported here shows that linguistic elements that drive perceptual simulation only do so in an utterance in which they have a literal, spatial meaning, suggesting that it is not just lexical associations but rather the construction of a model of the whole sentence’s meaning that drives simulation. And third, the experiments reported here show that spatial imagery is specific to the direction of motion—up or down—and not just the axis of motion, as previously demonstrated (Richardson et al., 2003). On the basis of these results, we argue for a view of lexical and sentential meaning in which words pair phonological form with specifications for imagery to be performed, and larger utterances compose these imagery specifications to drive a mental simulation of the content of the utterance.

Before looking in detail at the method used to address these issues in section 1.2., we provide an overview of work on mental simulation in language understanding in section 1.1.

76 1.1. *Mental simulation in language understanding*

77 To demonstrate the influence of language on mental imagery (we will be using “mental  
78 simulation” synonymously), it can be useful to consider the subjective experience of processing  
79 language associated with perceptual content. Answering questions like the following, for  
80 instance, may require mental imagery: What shape are a poodle’s ears? What color is the  
81 cover of *Cognitive Science*? Which is taller: a basketball hoop or a bus? (See also examples in  
82 Kosslyn, 1980). Critically, most people report that in answering such questions, they mentally  
83 picture or “look at” named objects; that they mentally rotate or otherwise manipulate these  
84 objects; that they are able to zoom in or out; and that they combine imagined objects in a single  
85 visual picture (Kosslyn, Ganis, & Thompson, 2001). These subjective visual experiences are  
86 triggered proximally by verbal input.

87 Mental imagery, then, can be defined as experience resembling perceptual or motor ex-  
88 perience occurring in the absence of the relevant external stimuli, in the case of perceptual  
89 experience; or without actual execution of motor actions, in the case of motor imagery. Im-  
90 agery has played a critical role in most theories of mind, starting at least as early as Aristotle.  
91 Modern investigations of imagery have demonstrated that it is integral to conceptual knowl-  
92 edge (Barsalou, Simmons, Barbey, & Wilson, 2003) and recall (Nyberg et al., 2001), can  
93 work unconsciously (Barsalou, 1999), can be used productively to form new configurations  
94 (Barsalou & Prinz, 1997), and works by activating neural structures overlapping with (or a Q6  
95 subset of) those used for perception and action (Ehrsson, Geyer, & Naito, 2003; Kosslyn et al.,  
96 2001).

97 Imagery has been argued in the literature on embodied cognition and especially cognitive  
98 linguistics to be critical to language. The shared central idea is that processing language  
99 activates internal representations of previously experienced events, or schematic abstractions  
100 over these (Lakoff, 1987; Langacker, 1991; Talmy, 2000). It is thus the (re)activation of modal Q7  
101 (e.g., perceptual or motor) content associated with particular described scenes that serves  
102 as the “engine” of meaning. This mental simulation process has been argued to be useful  
103 in the production of detailed inferences on the basis of language input (Narayanan, 1997),  
104 to prepare the understander for situated action (Bailey, 1997; Barsalou 1999; Glenberg &  
105 Kaschak, 2002), to build a situation model of a described scene (Zwaan, 1999), and to allow  
106 disambiguation (Bergen & Chang, 2005). In general, embodied approaches to language predict  
107 that understanding verbal input about events that can be perceived or performed will result in  
108 an individual’s tacit and automatic mental enactment of corresponding motor or perceptual  
109 imagery.

110 And this is precisely what has been observed in a number of recent studies. When processing  
111 language, understanders appear to activate imagery pertaining to the direction of motion of  
112 a described object (Glenberg & Kaschak, 2002; Kaschak et al., 2005), the shape (Stanfield  
113 & Zwaan, 2001), and the orientation (Zwaan et al., 2002) of described objects; the rate and  
114 length of (fictive) motion (Matlock, 2004b); the effector used to perform an action (Bergen  
115 et al., 2004; Bergen et al., 2003); and the axis (horizontal vs. vertical) along which action  
116 takes place (Lindsay, 2003; Richardson et al., 2003;).

117 In the remainder of this article, we concentrate on visual imagery evoked in response  
118 to natural language; in particular on the extent to which language triggers visual imagery

of motion or location in the upper or lower part of the visual field. Visual imagery lends itself well to empirical study because, as will be made clear in the next section, it is relatively easy to assess. Moreover, it is well-suited to the study of how language drives imagery because language that describes upward or downward motion or location occurs pervasively within languages. Because different classes of words like nouns (1a) and verbs (1b) have spatial meanings, we can study how these different word types contribute to the construction of a mental simulation. Spatial language is also advantageous because it tends to be multifunctional—language that describes literal, physical motion like 1b often also has figurative motion uses, where there is no literal motion of the described entity. Perhaps the most pervasive type of figurative motion is metaphorical motion (1c) in which an abstract event of some kind—in this case a change in quantity—is described with motion language. The multifunctionality of words denoting spatial motion allows us to investigate how the context of their use influences the manner in which words contribute to simulation:

- a. The *ground/roof* shook. 132
- b. The ant *climbed/dropped*. 133
- c. Stock prices *climbed/dropped*. 134

To develop a full account of how language drives mental imagery, we need to know what sorts of language (e.g., literal, figurative) result in what sorts of imagery, and what linguistic elements (e.g., nouns, verbs) trigger this imagery. The remainder of this section introduces the methodology used in this experiment and outlines previous work using this method.

### 1.2. Linguistic Perky effects 139

In a seminal study, Perky (1910) asked participants to imagine seeing an object (such as a banana or a leaf) while they were looking at a blank screen. At the same time, unbeknownst to them, an actual image of the same object was projected on the screen, starting below the threshold for conscious perception, but with progressively greater and greater illumination. Perky found that many participants continued to believe that they were still just imagining the stimulus and failed to recognize that there was actually a real, projected image even at levels where the projected image was perfectly perceptible to participants not simultaneously performing imagery.

Recent work on the Perky (1910) effect has shown that such interference of imagery on perception can arise not just from shared identity of a real and an imagined object, but also from shared location. Craver-Lemley and Arterberry (2001) presented participants with visual stimuli in the upper or lower half of their visual field while they were performing imagery either in the same region where the visual stimulus was or in a different region, or were performing no imagery at all. Participants were asked to say whether they saw the visual image, and were significantly less accurate at doing so when they were imagining an object (of whatever sort) in the same region than when they were performing no imagery or were performing imagery in a different part of the visual field.

A proposed explanation for these interference effects is that visual imagery makes use of the same neural resources recruited for actual vision (Kosslyn et al., 2001). In commonsense

159 terms, if a particular part of the retinotopically arranged visual system is being used for  
160 one function (say, imagery), then it will be significantly less efficient at performing another  
161 incompatible function (say, visual perception) at the same time. Interference of visual imagery  
162 on visual processing can be naturally used to investigate whether language processing also  
163 drives imagery. Rather than asking participants to imagine visual objects, experimenters can  
164 ask participants to process language hypothesized to evoke visual imagery of a particular  
165 type—of particular objects with particular properties or of objects in particular locations. If  
166 language of this sort selectively activates visual imagery, then we should expect a Perky-type  
167 effect that results in interference of the visual properties implied by the language on processing  
168 of displayed visual images.

169 This is precisely the tack taken by Richardson et al. (2003). In their work, participants  
170 first heard sentences whose content had implied spatial characteristics and then very quickly  
171 thereafter performed a visual categorization task (deciding whether a presented image on the  
172 screen was a circle or a square), where the location of an object they were asked to categorize  
173 could overlap with the imagery the sentence would supposedly evoke or not. The researchers  
174 reasoned that if sentence understanding entailed visual imagery, then there should be Perky-  
175 like interference on the object categorization task—that is, people should take longer to  
176 categorize an object when it had visual properties similar to the image evoked by the sentence.

177 Specifically, Richardson et al. (2003) suggested that processing language about concrete or  
178 abstract motion along different axes in the visual field (vertical vs. horizontal) leads language  
179 understanders to conscript the parts of their visual system that are normally used to perceive  
180 trajectories with those same orientations. For example, a sentence like 2a implies horizontal  
181 motion, whereas 2b implies vertical motion. If understanders selectively perform vertical or  
182 horizontal visual imagery in processing these sentences, then when they are asked immediately  
183 after presentation of the sentence to visually perceive an object that appears in their actual  
184 visual field, they should take longer to do so when it appears on the same axis as the motion  
185 implied by the sentence. Thus, after 2a (a horizontal-motion sentence), participants should  
186 take longer to categorize an object as a circle or a square when it appears to the right or left  
187 of the middle of the screen (on the horizontal axis) than it should take them to categorize an  
188 object when it appears above or below the middle of the screen (on the vertical axis):

- 189 a. The miner pushes the cart. [Horizontal]  
190 b. The ship sinks in the ocean. [Vertical]

191 An additional point of interest here concerns the nature of the sentences used. The ex-  
192 perimenters were interested in the spatial orientation not just of concrete verbs, like *push*  
193 and *sink*, but also abstract verbs, like *respect* and *tempt*. They wanted to determine whether  
194 abstract events, like concrete events, were selectively associated with particular spatial ori-  
195 entations. How abstract concepts are represented and understood is a critical question for all  
196 theories of meaning and understanding, but is particularly critical to simulation-based mod-  
197 els, which rely on perceptual and motor knowledge. There are insightful discussions of how  
198 abstract concepts can be grounded in embodied systems elsewhere (Barsalou, 1999; Barsalou  
199 & Wiemer-Hastings, 2005; Glenberg & Robertson, 2000; Lakoff, 1987), and the topic is  
200 explored in more depth in section 5.

Richardson et al. (2003) took verbs, with associated horizontality–verticality and concreteness–abstractness ratings determined through a norming study (Richardson et al., 2001), and presented them to participants in the interest of ascertaining whether they would induce Perky-like effects on the categorization of visual objects (shapes). These objects were presented on the screen in locations that overlapped with the sentences’ implied orientation. After seeing a fixation cross for 1 sec, participants heard a sentence; then, after a brief pause (randomly selected for each trial from among 50, 100, 150, or 200 msec), they saw a visual object that was either a circle or a square positioned in one of the four locations on the screen (right, left, top, or bottom). Their task was to press a button indicating the identity of the object (1 button each for “circle” and “square”) as quickly as possible:

- a. The miner pushes the cart. [Concrete Horizontal] 211
- b. The plane bombs the city. [Concrete Vertical] 212
- c. The husband argues with the wife. [Abstract Horizontal] 213
- d. The storeowner increases the price. [Abstract Vertical] 214

Q8 The results indicated a clear interference effect—participants took longer to categorize objects on the vertical axis after vertical sentences (as compared with vs. horizontal sentences), and vice versa for objects on the horizontal axis. Intriguingly, *post hoc* tests (which the authors explicitly indicated were, strictly speaking, statistically unwarranted) showed that this interference effect was significant for abstract sentences but not for the concrete sentences (see section 6 for details).

Q9 It is important to underline at this point that the expected (and observed) effect was interference between language and visual perception using the same part of the visual field. This contrasts with other work (Glenberg & Kaschak, 2002; Zwaan et al., 2002), which has found facilitatory compatibility effects. Briefly, it appears that when the same cognitive resources are used for two tasks at the same time, as is believed to occur with the very short latency between sentence and object perception in the Perky (1910) task (50–200 msec), we observe interference. The explanation for this interference is that the same cognitive resources cannot be adequately used to perform two distinct tasks at the same time. It should be difficult then for a participant to use a particular part of their visual system to simultaneously imagine an object in a particular location in the imagined visual field and also perceive a distinct object in the same location of their real visual field if the two processes use the same parts of the visual system—the claim at the heart of the visual imagery hypothesis. By contrast, when there is enough time between the tasks for priming to take place, such as the 250 msec or more in studies like Glenberg and Kaschak (2002), Stanfield and Zwaan (2001), and Zwaan et al. (2002), facilitation is observed (Bergen, in press ; Kaschak et al., 2005).

Although the work reported by Richardson et al. (2003) provided key insights into the relationship between imagery and language, it also leaves several questions unanswered; questions that we will explore in this article. First, why would abstract sentences but not literal sentences generate the expected Perky (1910) effect? No simulation-based account of language understanding, nor any account of language understanding that we are aware of, would predict that abstract but not literal spatial language should yield perceptual imagery.

242 Second, Richardson et al.'s (2003) study was not designed to tell us what linguistic elements  
243 in the sentences were yielding the observed effects. The sentences used different argument  
244 structures including both transitive and intransitive structures and had subjects and objects  
245 whose own vertical or horizontal associations were not controlled for.

246 Third, when one takes a close look at the sentences appearing in the abstract condition,  
247 their verbs fall into varied semantic classes. The abstract category includes relatively abstract  
248 verbs like *hope* and *increase* as well as relatively concrete ones like *argue* and *give*. Moreover,  
249 with few exceptions, the nouns used in the sentences are almost entirely concrete, denoting  
250 people, physical objects, and places. As a result, it may be that even abstract verbs, when  
251 combined with concrete arguments, evoke imagery of concrete situations. For instance, the  
252 abstract horizontal sentence, "The husband argues with the wife," might well yield imagery  
253 of a scene in which the two participants in the argument are arrayed horizontally, in the way  
254 that two people normally would when arguing. As a result, the question remains open what  
255 types of "abstract" verbs, combined with what types of arguments into abstract sentences,  
256 yield spatial imagery.

257 Fourth and finally, this experiment intentionally conflated the up and down positions and  
258 the right and left positions. For example, both sentences in the following list (4) are in  
259 the Concrete Vertical condition, despite the fact that they describe movement in opposite  
260 directions. Although it could be that the entire imagined vertical axis is used to process both  
261 of these sentences, the absence of any significant effect for concrete sentences in Richardson  
262 et al.'s (2003) study suggests that there may be something more complicated going on. It could  
263 be instead that sentences describing downwards motion, like 4a, yield spatial processing in  
264 the lower part of the imagined visual field; whereas upward sentences, like 4b, do the same  
265 in the upper part of the imagined visual field. If so, then subsets of the stimuli in each of the  
266 concrete conditions would actually have imagery and objects in different parts of the visual  
267 field:

- 268 a. The ship sinks in the ocean.  
269 b. The strongman lifts the barbell.

270 Thus, the current state of affairs still leaves open the three questions identified earlier.  
271 Namely, (a) what linguistic cues trigger mental simulation, (b) what sorts of language (lit-  
272 eral, metaphorical, abstract) result in mental simulation, and (c) how detailed is the mental  
273 simulation?

## 274 **2. Experiment 1: upward and downward motion**

275 Does language denoting literal motion in a particular direction drive visual imagery local-  
276 ized to the same part of the visual field? Our first experiment followed Richardson et al. (2003)  
277 but aimed to answer the outstanding questions of what linguistic elements drive simulation and  
278 how detailed it is. The design here controlled for the linguistic components of sentences and  
279 separated the vertical axis into distinct up and down regions. Based on prior work showing that  
280 the Perky (1910) effect is location specific (Craver-Lemley & Arterberry, 2001), we expected

that people would take longer to identify objects in the upper or lower part of the visual field 281  
following sentences denoting scenes that canonically take place in the same locations. 282

To reduce the range of possible linguistic factors influencing imagery, we used bare in- 283  
transitive sentences (sentences with only a subject noun phrase and a main verb). The verbs, 284  
as determined by a norming task, all denoted literal motion in a particular direction. This 285  
meant that only upward and downward motion could be used, as there are no verbs in English 286  
that denote rightward or leftward motion. All subject nouns in the critical sentences were 287  
determined through a norming study to be unassociated with *upness* or *downness*. Critical 288  
sentences thus fell into two directional conditions (up and down): 289

- a. The mule climbed. [Upward motion] 290
- b. The chair toppled. [Downward motion] 291

### 2.1. Method 292

Sixty-five native speakers of English participated in exchange for course credit in an 293  
introductory linguistics class at the University of Hawaii. 294

Participants wore headphones and sat in front of a computer screen. They heard sentences 295  
and looked at geometric shapes that were presented in one of four locations on the screen. 296  
They were instructed to quickly press one of two buttons to identify whether the shape was a 297  
square (by pressing “x”) or a circle (by pressing “z”). Each trial began with a fixation cross 298  
that appeared in the middle of the screen for 1,000 msec. Next, a sentence was presented 299  
auditorily, followed by an ISI of 200 msec (during which time the screen was blank). Then a 300  
circle or a square appeared in the top, bottom, left, or right part of the screen for 200 msec. All 301  
objects appeared the same distance from the fixation cross at the center of the screen, along a 302  
central axis (e.g., objects in the upper part appeared directly over the fixation cross). 303

Q10

In critical trials, sentences denoted either upward motion or downward motion (5), and 304  
the object appeared in the upper or lower region. Filler trials were randomly interspersed. 305  
Some filler trials included a short yes–no comprehension question to ensure that participants 306  
attended to the meaning of the sentences. For instance, the filler sentence, “The branch split,” 307  
was followed by the question, “Did the branch break?” Filler trials included as many up- and 308  
down-related sentences as appeared in the critical trials, but all of these were followed by an 309  
object on the left or right—all of these sentences were selected from among the sentences 310  
discarded through the norming study. 311

The constraints imposed by this design, that only intransitive verbs denoting upward or 312  
downward motion could be used, translated into a relatively small number of candidate verbs. 313  
In English, there are only 5 to 10 verbs denoting either upward or downward motion. Because 314  
of the small number of possible verbs of each type, the entire list of sentences was presented 315  
twice to each participant—once followed by a shape in the upper region and once followed 316  
by a shape in the lower region of the screen. To ensure that there was distance between the 317  
two instantiations of each critical sentence, the experiment was broken into two halves, each 318  
of which contained all critical sentences in a random order. The order of the two halves was 319  
manipulated to create two lists. Participants were randomly assigned to one of these lists. 320



321 2.2. Norming

322 In constructing stimuli, we conducted a norming study to ensure that the critical sentences  
 323 had several properties. For each type of sentence, we aimed to include sentences in the up  
 324 condition that were no more or less meaningful than sentences in the down condition, and to  
 325 have as little difference as possible in processing time between the two groups of sentences.  
 326 Second, and more critically, we wanted to ensure that the sentences, which had only a subject  
 327 and a verb, differed in terms of their upness or downness only because of one manipulated  
 328 word. Therefore, the sentential subjects used in the critical sentences in this experiment  
 329 were constrained to be equally neutral for their up–down associations (e.g., *chair* and *donkey*),  
 330 whereas the verbs denoted significantly different up/down meanings (e.g., *climb* and *descend*).

331 A total of 57 native speakers of English from the University of Hawaii community partic-  
 332 ipated in the norming study in exchange for credit in an introductory linguistics class. They  
 333 performed three tasks. First, they completed a sentence reading task in which sentences were  
 334 presented and participants were instructed to press a button as soon as they understood the  
 335 meaning of the sentence. They were then asked to rate the meaningfulness of the sentence on  
 336 a scale ranging from 1 (*least meaningful*) to 7 (*most meaningful*). Next they were given a list  
 337 of words, either nouns or verbs, and were asked to rate them as to how strongly their meanings  
 338 were associated with up or down—1 (*the least up- or down-associated*) to 7 (*the most up- or*  
 339 *down-associated*). One group of participants rated only upness, the other only downness.

340 The critical stimuli in the upness or downness rating task included verbs that the exper-  
 341 imenters hypothesized to denote motion events canonically moving upward or downward  
 342 and nouns denoting objects canonically located above or below an observer’s head, and the  
 343 sentences in the reading and meaningfulness part of the norming study were constructed from  
 344 these words. In addition, each group of participants saw one half of the proposed filler sen-  
 345 tences, which were expected to be meaningful; and the other half with the verbs and participant  
 346 nouns randomized across sentences, which were thus unlikely to be meaningful. Finally, each  
 347 participant saw 15 sentences with transitive verbs used intransitively, which were also unlikely  
 348 to be judged meaningful.

349 One participant was removed from the norming study analysis for having a mean reac-  
 350 tion time (RT) more than 2 *SDs* greater than the grand mean. We also removed all trials  
 351 with RTs less than 350 msec, as these sentences were unlikely to have been thoroughly  
 352 understood.

353 In selecting sentences for the main experiment, we eliminated all sentences with extremely  
 354 fast or slow RTs, low meaningfulness ratings, nouns with strong up or down associations,  
 355 or verbs without strong up or down associations. This left five sentences in each critical  
 356 condition.<sup>1</sup> The mean upness and downness ratings for the nouns selected for the main study  
 357 are shown in Table 1. The nouns in the upward motion sentences were not significantly more  
 358 up-related than those in downward motion sentences:  $F(1, 28) = 0.55, p = .47$ ; nor were  
 359 they significantly more down-related (although the effect here approached significance),  $F(1,$   
 360  $27) = 3.56, p = .07$ . Turning to the verbs, it was crucial that the verbs used in two conditions  
 361 differed from each other in terms of their upness and downness. Overall, verbs were classified  
 362 as expected: The verbs in the two literal conditions differed significantly in their upness ratings,  
 363  $F(1, 28) = 117.65, p < .001$ ; and their downness ratings,  $F(1, 27) = 134.54, p < .001$ .

Table 1  
 Results of norming studies in which participants, on a 7-point scale, rated nouns and verbs on upness and downness

Experiments	Nouns				Verbs			
	Up Avg	SD	Down Avg	SD	Up Avg	SD	Down Avg	SD
Experiment 1								
Down (Verb)	2.04	1.65	2.31	1.82	1.85	1.09	5.39	1.16
Up (Verb)	2.12	1.76	2.00	1.48	5.18	1.43	2.35	1.40
Experiment 2								
Down (Noun)	1.99	1.72	4.61	2.18	2.14	1.31	2.06	1.35
Up (Noun)	5.37	1.91	2.09	1.62	2.19	1.41	2.04	1.16
Experiment 3								
Down (Metaphor)	4.64	2.00	4.33	2.14	1.85	1.09	5.39	1.16
Up (Metaphor)	4.45	2.01	4.34	2.09	5.18	1.43	2.35	1.40
Experiment 4								
Down (Abstract)	4.35	2.30	4.05	2.19	1.63	0.82	4.40	1.32
Up (Abstract)	4.37	2.09	4.10	2.16	4.52	1.75	1.54	0.79

Note.  $n = 28$ . Avg = average.

Also of interest are the mean reading times and meaningfulness ratings, shown in Table 2. 364  
 Repeated-measures analyses of variance (ANOVAs) revealed a reliable difference in reading 365  
 times,  $F(1, 28) = 12.39, p < .01$ ; and a marginally significant difference in meaningfulness, 366  
 $F(1, 28) = 4.10, p = .05$ . Although it is certainly not ideal to have such differences between 367  
 conditions, it was a necessary artifact of the design, as very few verbs exist in English that can 368  
 denote intransitive upward motion. This will be discussed in more detail in the next section. 369

Table 2  
 Results of norming studies in which participants read sentences and rated them on 7-point scale of  
 meaningfulness

Experiments	Reaction Time		Meaningfulness	
	M	SD	M	SD
Experiment 1				
Down (Verb)	1,515	631	6.16	0.81
Up (Verb)	1,844	813	5.81	0.96
Experiment 2				
Down (Noun)	1,691	828	6.31	0.88
Up (Noun)	1,554	624	6.48	0.88
Experiment 3				
Down (Metaphor)	1,970	832	5.41	1.04
Up (Metaphor)	2,011	1,036	5.59	0.92
Experiment 4				
Down (Abstract)	1,932	875	6.13	0.80
Up (Abstract)	1,811	806	6.12	0.75

Note.  $n = 28$ .

Table 3  
 Mean reaction time (RT) in milliseconds for object categorization in upper and lower quadrants of the screen

Experiments	Object in Lower Quadrant			Object in Upper Quadrant		
	Mean RT	SD	SE	Mean RT	SD	SE
Experiment 1						
Down (Verb)	551	255	32	542	240	30
Up (Verb)	526	205	26	603	270	34
Difference (msec)	+25			-61		
Experiment 2						
Down (Noun)	550	221	28	506	245	20
Up (Noun)	508	218	30	526	247	22
Difference (msec)	+42			-20		
Experiment 3						
Down (Metaphor)	516	283	23	532	228	25
Up (Metaphor)	535	235	24	531	240	24
Difference (msec)	-19			+1		
Experiment 4						
Down (Abstract)	589	230	29	575	222	28
Up (Abstract)	593	268	33	600	317	40
Difference (msec)	-4			-25		

370 2.3. Results

371 Only participants who answered the sentence comprehension questions with at least 85%  
 372 accuracy were included in the analysis—this eliminated 1 participant. Another participant  
 373 was excluded for answering the object categorization questions with only 79% accuracy.  
 374 None of the remaining participants performed at less than 90% accuracy on the critical trials.  
 375 Responses that were 3SDs above or below the mean for each participant were removed and  
 376 replaced with values 3 SDs above or below the mean for that participant.<sup>2</sup> This resulted in  
 377 changes to less than 1% of the data.

378 The mean RTs for the literal sentences displayed in the first two data rows of Table 3  
 379 show a clear interaction effect of the predicted kind. Objects in the upper part of the visual  
 380 field are categorized faster following literal down sentences than they are following literal  
 381 up sentences, and the reverse is true for visual objects in the lower part of the visual field  
 382 (although this latter effect does not appear to be as strong). A repeated-measures ANOVA by  
 383 participants showed the predicted interference effect through a significant interaction between  
 384 sentence direction (up or down) and object location (up or down),  $F(1, 63) = 5.03, p < .05$ ;  
 385 partial  $\eta^2 = 0.07$ ). There were no significant main effects of sentence type or object location.  
 386 With only five items in each condition, it would be unrealistic to expect an ANOVA using  
 387 items as a random factor to show significance. Moreover, because the set of stimuli in each  
 388 condition effectively constitutes the population of relevant items, and are not random samples  
 389 from that population, it would not make sense to perform such an analysis in any case. As  
 390 shown in Table 4, however, all up sentences had longer RTs in the Up Object condition than in  
 391 the Down Object condition (by at least 30 msec), suggesting that the interference effect holds

Table 4  
Mean reaction time in milliseconds for object categorization in upper and lower quadrants of the screen for Up and Down sentences in Experiment 1, by sentence

Sentences	Object Up	Object Down
Up		
The Cork Rocketed.	645	458
The Mule Climbed.	529	493
The Patient Rose.	635	591
The Lizard Ascended.	644	541
The Dolphin Soared.	570	539
Down		
The Glass Fell.	514	611
The Chair Toppled.	605	625
The Cat Descended.	399	578
The Pipe Dropped.	588	492
The Stone Sank.	614	456

for all the Literal Up sentences. Similarly indicative of interference, three out of five of the 392  
Literal Down sentences had longer RTs in the Down than in the Up condition. Looking at the 393  
items individually, it seems that the interference effect is stronger with Literal Up sentences, 394  
which yielded much slower response times on average on objects in the upper position than 395  
those in the lower position. 396

To deal with the problem of a small set of potential verbs, the design of this study pre- 397  
sented each critical sentence once with the visual stimulus in the upper region and once 398  
with the visual stimulus in the lower region. Because the repetition of stimuli runs the risk 399  
of inducing carryover effects (e.g., participants develop different strategies for responding 400  
to stimuli they have or have not seen already), we performed a *post hoc* analysis to de- 401  
termine whether such effects accounted for the results reported here. To do this, we an- 402  
alyzed the data from the first half of the experiment only, which included just the first 403  
presentation of each sentence. The results, seen in Table 5, are not statistically signifi- 404  
cant,  $F(1, 63) < 1$ , as might be expected given the low number of stimuli per condition 405  
per participant (2.5). However, the trend is in same direction as the full results, suggesting 406  
that carryover effects were not responsible for the critical Perky-like interference effect we 407  
observed. 408

Table 5  
Mean reaction time (RT) in milliseconds for object categorization in upper and lower quadrants of the screen, for the first half of Experiment 1 only

Category	Object in Lower Quadrant		Object in Upper Quadrant	
	Mean RT	SE	Mean RT	SE
First Half Only				
Down (Verb)	604	34	561	28
Up (Verb)	593	29	626	39
Difference (RT)	+11	-65		

409 *2.4. Discussion*

410 The significant interaction effect observed here with sentences denoting upward or down-  
411 ward motion leads to two important conclusions. The first involves the specificity of the  
412 imagery associated with these sentences. Although it is known (Richardson et al., 2003) that  
413 the axis of motion of a sentence is accessed during language processing, this study provides  
414 evidence that the spatial grain of visual imagery is in fact even more detailed than this. Because  
415 sentences denoting upward and downward motion selectively interfered with categorizing ob-  
416 jects in the same part of the visual field, we can see that motion imagery in response to these  
417 sentences is specific to the location in which the content of the utterance would take place,  
418 not just the axis.

419 Second, unlike the post hoc report on Richardson et al.'s (2003) results, we observed a  
420 reliable interaction with concrete sentences denoting physical motion. This finding is more  
421 squarely in line with what is predicted by theories of perceptual simulation in language  
422 understanding—that literal language about space should be processed using those neurocog-  
423 nitive systems responsible for perceiving the same aspects of space. As suggested in the intro-  
424 duction, these results suggest that the lack of an effect for concrete sentences in Richardson  
425 et al. may have resulted from the conflation of the up and down directions into a single level.  
426 As we have seen here, sentences denoting upward motion result in interference in the upper  
427 part of the visual field. Thus, it would not be not surprising if, when upward- and downward-  
428 oriented sentences are combined in a single condition, their effects cancelled each other  
429 out.

430 The effect we observed here was especially strong for sentences denoting upward motion.  
431 Why might upward motion sentences show this effect a stronger effect than downward motion  
432 sentences? One plausible explanation is that the difference results from the slightly (although  
433 not significantly) greater time it took participants to process the upward motion sentences.  
434 Perhaps they had not completed the comprehension process at the point in time when the  
435 visual object was presented—in this case, continued sentences imagery would yield a greater  
436 interference effect.

437 Another possible explanation points to differences in the likelihood of the two types of  
438 events described. In everyday life, we often observe objects moving downward, even when  
439 there is no force acting on them. By contrast, we rarely observe objects moving upward,  
440 especially without force overtly exerted on them. Because upward motion events without an  
441 external agent are less common in the world than equivalent downward events, individuals  
442 might have a need for greater simulation (more time, more effort) in the case of upward motion.  
443 This would result in greater interference with visually categorizing objects in the upper part  
444 of the visual field.

445 Regardless of the details of this effect, the crucial manipulation that yielded it was the use  
446 of verbs that were strongly associated with upward or downward motion. From the simulation-  
447 based perspective, the effects are perfectly predictable because verbs of motion are supposed to  
448 indicate processes and relations holding of entities. What would happen, though, if nouns were  
449 manipulated while verbs were held constant? Do nouns denoting objects that are canonically  
450 associated with the upper or lower part of the visual field yield the same sort of interference?  
451 This is the topic of the next study.

### 3. Experiment 2: up- or down-associated nouns

452

In Experiment 1, we found a significant interference effect when a motion verb in a sentence denoted movement in a particular direction and a visual object that was subsequently categorized appeared in the same part of the visual field. In this study, we investigate whether the same effect can be produced by manipulating the subject noun alone.

Recent work on visual imagery during language understanding has demonstrated that mentioned objects are represented with a good deal of visual detail. In work in a paradigm different from the current one, Stanfield and Zwaan (2001) and Zwaan et al. (2002) had participants read sentences then name or make a judgment about an image of an object that had been mentioned in the sentence. They found that implied orientation of objects in sentences like the following (6) affected how long it took participants to perform the object judgment task. Participants took longer to respond to an image that was incompatible with the implied orientation or shape of a mentioned object. For example, reading a sentence about a nail hammered into a wall primed the horizontal nail image, as contrasted with a sentence about a nail hammered into the floor. Similar results were found for shape of objects, such as a whole egg versus a cracked egg in a pan. These results imply that shape and orientation of objects are represented in mental imagery during language understanding:

- a. The man hammered the nail into the floor. 469
- b. The man hammered the nail into the wall. 470

People also seem to mentally represent the locations of objects in space. Eye-tracking evidence from narrative comprehension shows that listeners looking at a blank screen tend to look at those locations in space where mentioned objects and events would appear both during comprehension (Spivey & Geng, 2001) and recall (Johansson, Holsanova, & Holmqvist, 2005). These studies, along with earlier work on mental models (e.g., Bower & Morrow, 1990), show that when objects are described as appearing in particular locations, this spatial location is represented in an analogue fashion. However, it is not yet known whether the location where an object is canonically found (e.g., above or below an observer) is automatically engaged as part of the mental simulation evoked by an utterance.

The question of whether nouns that denote objects which happen to be canonically located in up or down locations can yield perceptual interference effects is crucial to understanding what factors make an utterance likely to produce visual simulations with particular properties. If nouns themselves can trigger imagery in the upper or lower part of the visual field, then this could potentially help to explain some of the effects reported by Richardson et al. (2003).

#### 3.1. Method

485

A total of 63 students from the same population described in Experiment 1 (who had not participated in Experiment 1) participated in this study. The method was globally identical to that in Experiment 1, with the exception of the critical sentences. In this experiment, participants listened to critical sentences whose subject nouns were canonically associated with upness or downness and whose verbs were vertically neutral (no upness or downness)—

490

491 for example, “The cellar flooded,” and “The ceiling cracked.” The sentences were constructed  
492 from items selected from the norming study described in Experiment 1. In the norming study,  
493 the Up and Down sentences showed no significant difference in RT:  $F(1, 27) = 0.89$ ,  $p =$   
494  $.35$ ; or in meaningfulness:  $F(1, 27) = 2.60$ ,  $p = .12$  (see Table 2).

495 Moreover, the verbs in the two noun conditions did not differ significantly in either their  
496 upness ratings,  $F(1, 28) = 0.13$ ,  $p = .72$ ; or their downness ratings,  $F(1, 27) = 0.01$ ,  $p = .93$   
497 (see Table 1). By contrast, the nouns in the up versus down sentences were highly differentiated  
498 in terms of upness:  $F(1, 28) = 215.16$ ,  $p < .001$ ; and down-ness:  $F(1, 27) = 132.31$ ,  $p <$   
499  $.001$ . These norming results serve to ensure that any interference effects observed on the object  
500 categorization task would result from the differences in the up or down associations of nouns  
501 alone, not in differences between the verbs.

### 502 3.2. Results

503 Response times from two participants whose mean response times fell 2 *SDs* above the  
504 mean for all participants were removed. In addition, response times for two other participants  
505 were removed for answering the comprehension questions with less than 80% accuracy. In  
506 the remaining data set, responses more than 3 *SDs* from each participant’s mean RT were  
507 replaced with values 3 *SDs* from their mean. This resulted in the modification of less than 1%  
508 of the data.

509 Considering only correct responses, the means were as shown in Table 3. As with the  
510 verb manipulation in Experiment 1, there was interference in the predicted direction between  
511 sentence direction and object location. Indeed, a repeated-measures ANOVA by participants  
512 showed a significant interaction between object location and sentence direction,  $F(1, 58) =$   
513  $5.76$ ,  $p < .05$ ; partial  $\eta^2 = 0.09$ . There were no significant main effects of object location or  
514 sentence direction. Again, there were too few items to expect an item analysis using ANOVA  
515 to yield significant results, but looking at them individually (Table 6), we see that almost  
516 all of the sentences with down-associated subject nouns yielded faster categorization when  
517 the subsequent object appeared in the upper part of the visual field. It is interesting to note  
518 that the one exceptional sentence in this group, “The submarine fired,” might be construed  
519 as encoding upward movement—that is, when submarines fire ballistic missiles rather than  
520 torpedoes, they typically fire upward. The sentences with up-related subject nouns showed  
521 the opposite tendency, as predicted. Namely, the majority yielded faster response times to the  
522 categorization task when the object appeared in the lower part of the screen.

### 523 3.3. Discussion

524 The striking finding from this study is that sentences with subject nouns that are canonically  
525 associated with upness or downness selectively interfere with the visual processing of objects  
526 in the same parts of the visual field. This is in line with other work on visual imagery associated  
527 with objects in sentence understanding, which shows that both the shape (Stanfield & Zwaan,  
528 2001) and orientation (Zwaan et al., 2002) of objects are primed by sentences that imply those  
529 particular shapes or orientations for objects.

Table 6  
Mean reaction time in milliseconds for object categorization in upper and lower quadrants of the screen for Up and Down sentences in Experiment 2, by sentence

Sentences	Object Up	Object Down
Noun Down		
The Cellar Flooded.	478	511
The Grass Glistened.	515	568
The Ground Shook.	533	708
The Shoe Smelled.	457	484
The Submarine Fired.	547	474
Noun Up		
The Ceiling Cracked.	515	486
The Rainbow Faded.	592	412
The Roof Creaked.	538	609
The Sky Darkened.	506	472
The Tree Swayed.	479	561

Note that unlike the sentences with verbs denoting upward or downward motion described in Experiment 1, the sentences with up- or down-associated nouns did not display an asymmetry between a strong effect in up sentences and a small effect in down sentences. This would tend to support either of the explanations given there—that this asymmetry in Experiment 1 was due to either a difference in processing times between the sentences (which was not seen in the norming data for the sentences in Experiment 2), or that it arose due to the unusualness of intransitive motion (because the sentences in Experiment 2 did not encode upward or downward motion so much as up or down location). Either of these accounts would predict the asymmetry to disappear in this second study. In agreement with this prediction, we can see that the effect is not stronger for up sentences than down ones—in fact, the tendency seems to be weakly in the opposite direction.

Further, it is worth noting that the interference effect was observed in both Experiments 1 and 2, despite substantial differences between them. Sentences in Experiment 1 (e.g., The mule climbed) denoted dynamic motion events, whereas in Experiment 2 sentences (e.g., The grass glistened) described a static object canonically found in a particular location. We might expect to find a greater interference effect for the first experiment if a sentence denoting motion was paired with motion of an incompatible object observed on the screen, and work in such a vein has shown compatibility effects of apparent motion toward or away from the participant (Zwaan, Madden, Yaxley, & Aveyard, 2004). An additional difference between the experiments involved whether the upness or downness of the sentence was carried by the noun or verb, grammatical classes that have been noted (Kersten, 1998) to be differently associated with motion. And yet, the two studies showed the same global interference effect, suggesting that it is a matter of the interpretation of the scene described by the sentences as a whole, rather than the contributions of individual words in the sentence, that drives the interference.

Despite the reliability of the interference effect shown in these first two studies, we have not conclusively shown yet that the mental imagery is driven by the processing of an entire sentence. The effects we have observed so far could instead result from some sort of



557 strictly lexical process. Perhaps the lexical representations for words like *ceiling* and *rise*  
558 share a common feature [+UP], and it is this feature, rather than a dynamic simulation of  
559 the utterance's content, that is causing the interference effects. Granted, one might be more  
560 likely to anticipate facilitatory priming on this lexical semantic feature account, but because  
561 inhibitory lexical effects are also observed in certain cases, and to eliminate the possibility that  
562 the effect is simply lexical, a third experiment used the same set of verbs described in the first  
563 study but with subject nouns that could not literally move up or down. Finding no interaction  
564 effect here would suggest that the interference was a result of sentence interpretation and not  
565 simply lexical semantics.

#### 566 4. Experiment 3: metaphorical motion

567 Language about motion in a direction, or about objects located in a given location, yielded  
568 significant interference on a visual perception task in the first two studies. To investigate  
569 whether this effect was the result of lexical or sentential interpretation, we performed a third  
570 experiment testing whether sentences that included motion verbs but did not denote literal  
571 motion would also interfere with object categorization.

572 Verbs of motion can be used cross-linguistically to describe events that do not involve literal  
573 motion, such as fictive motion (7a and 7b; Matlock, 2004a; Talmy, 2000) and metaphorical  
574 motion (7c and 7d; Lakoff & Johnson, 1980):

- 575 a. The drainpipe *climbs* up the back wall of the house.
- 576 b. Starting at the house, the fence *drops* down quickly to the ocean.
- 577 c. Oil prices *climbed* above \$51 per barrel.
- 578 d. Mortgage rates *dropped* further below 6 percent this week.

579 The interpretation processes involved in understanding figurative language have been a  
580 matter of significant research and debate. Some work has demonstrated that language users  
581 access internal representations of space and motion when performing reasoning tasks about  
582 abstract concepts understood metaphorically in terms of these concrete notions (Boroditsky,  
583 2000; Boroditsky & Ramscar, 2002; Gibbs, Bogdonovich, Sykes, & Barr, 1997). Moreover,  
584 there is limited evidence that processing connected discourse using metaphor proceeds most  
585 quickly when conventional metaphorical expressions are used (Langston, 2002). However,  
586 we do not yet know whether simply processing metaphorical motion language makes use of  
587 spatial representations. Critically, if the effect observed above in the first two experiments is  
588 simply lexical or if figurative language yields the same visual imagery that literal language  
589 does, then we should expect to see no difference when the same experiment described above  
590 is conducted with figurative upward or downward motion sentences rather than literal ones.  
591 However, if the effect observed in the previous experiments is due to the interpretation of the  
592 sentence—where a participant mentally simulates the described scene—and does not simply  
593 result from the lexical semantics of constituent words (and if figurative language differs in  
594 some ways from literal language interpretation), then we expect to see a significant decrease  
595 in the interference effect with metaphorical sentences. In the most convincing scenario, we

would observe the significant interference effect triggered by literal sentences to disappear 596  
with figurative ones. 597

4.1. Method 598

All the motion verbs used in the first study on literal sentences (section 2) can also be used 599  
to describe changes in quantity or value of entities that do not have physical height, such as oil 600  
prices or mortgage rates (7c and 7d). Thus, to create metaphorical sentences, we used subjects 601  
such as *rates* and *prices* along with the same motion verbs used in the first experiment to 602  
produce metaphorical sentences. The sentences were normed as described in section 2.2. The 603  
up and down metaphorical sentences showed no significant difference in RT,  $F(1, 27) = 0.07$ , 604  
 $p = .79$ ; or in meaningfulness,  $F(1, 27) = 0.97$ ,  $p = .33$  (Table 2). The nouns in metaphorical 605  
up versus down sentences were not rated differently in upness:  $F(1, 28) = 1.21$ ,  $p = .28$ ; or 606  
in downness:  $F(1, 27) = 0.003$ ,  $p = .95$ ; whereas the verbs were, as seen in Table 1. 607

In all respects other than the critical stimuli, the experiment was exactly as described earlier, 608  
and was in fact run together with Experiment 2. 609

4.2. Results 610

As can be seen from Table 3, by contrast with the literal verb and noun sentences, there was 611  
no significant interaction between sentence direction and object location with the metaphorical 612  
sentences,  $F(1, 58) = 0.43$ ,  $p = .52$ ; partial  $\eta^2 = 0.01$ ; nor were there significant main 613  
effects of object location or sentence direction. The analysis of items (Table 7) reveals the 614  
same absence of interference: More sentences in the down condition yielded faster response 615  
times when the object was in the lower half of the visual field, and the reverse was true for 616  
metaphorical up sentences. Both of these tendencies were the reverse of the predicted direction 617  
of the Perky (1910) effect. 618

Table 7  
Mean reaction time in milliseconds for object categorization in upper and lower quadrants  
of the screen for Up and Down sentences in Experiment 3, by sentence

Sentences	Object Up	Object Down
Metaphorical Down		
The Market Sank.	576	478
The Percentage Dropped.	570	518
The Quantity Fell.	491	490
The Rates Toppled.	473	493
The Ratio Descended.	548	600
Metaphorical Up		
The Amount Rose.	494	601
The Cost Climbed.	581	482
The Fees Ascended.	568	476
The Numbers Rocketed.	517	593
The Rating Soared.	492	523

619 *4.3. Discussion*

620 The absence of an interference effect in the metaphorical sentences confirms that the effects  
621 observed in Experiments 1 and 2 were the result of sentence interpretation and not just of  
622 the activation of lexical semantics. The verbs in Experiments 1 (literal motion sentences)  
623 and 3 (metaphorical sentences) were the same, and the subject nouns in the two sentence  
624 conditions in each experiment had identical up-down ratings. Consequently, the presence of  
625 interference effects in the literal sentences must result from understanding processes applied  
626 to the sentences as a whole.

627 A second notable finding here is that metaphorical sentences are not processed the same  
628 way as their literal counterparts with respect to visual imagery. This is initially surprising  
629 because many studies have shown that a literal source domain is in fact activated during  
630 the processing of metaphorical language (Boroditsky, 2000; Boroditsky & Ramscar, 2002;  
631 Gibbs et al., 1997). However, these results are not inconsistent because all that the cur-  
632 rent study indicates is that metaphorical and literal motion language differ in terms of their  
633 use of visual imagery at a particular point in time during sentence comprehension. It is  
634 possible that the sentences used would in fact trigger visual imagery, just with a different  
635 time course; or, for that matter, different intensity or variability than the literal language.  
636 One obvious avenue of research would be to apply eye-tracking techniques used for the  
637 closely related case of fictive motion (e.g., *The road runs through the woods*; Matlock  
638 & Richardson, 2004; Richardson & Matlock, in press) to metaphorical language like the Q11  
639 sentences used in this experiment. However, we must leave this question open for further  
640 investigation.

641 The results from the first two experiments suggest that literal sentences of different types  
642 give rise to visual imagery. Therefore, we turn to the question of abstract motion sentences.  
643 Richardson et al. (2003) reported a significant interference effect for abstract sentences but  
644 none for concrete sentences. By contrast, as we have seen, the current study (which differed  
645 in terms of the composition of the sentences and the manipulation of the spatial dimension)  
646 did yield interference with literal sentences. What is the relation between the visual imagery  
647 performed for literal and abstract motion language?

648 **5. Experiment 4: abstract verbs**

649 This experiment tested whether abstract sentences produce location-specific interference  
650 on a visual categorization task. Our abstract sentences, like the metaphorical sentences in  
651 Experiment 3, denoted changes in quantity but did so using verbs that did not also have a  
652 concrete meaning denoting change in height (verbs such as *increase* and *wane*). Embodied  
653 accounts of conceptual representation and language understanding (Barsalou, 1999; Glenberg  
654 & Robertson, 2000; Lakoff, 1987) argue that all concepts, whether concrete or abstract,  
655 are ultimately grounded in terms of embodied individual human experience in the world.  
656 The grounding of concrete concepts can be straightforwardly accounted for in terms of the  
657 perceptual, motor, and perhaps even affective content of experiences an agent has when dealing  
658 with instances of them. Indeed the evidence from the first two experiments in the current work

indicates that understanding language about motion in a particular direction or about an object canonically located in a particular place involves accessing the perceptual correlates of perceiving the described scene. It might similarly be argued that abstract concepts like changes in quantity or value can be grounded in terms of changes in physical location. This is precisely what is suggested by Richardson et al.'s (2003) finding that abstract sentences yield interference on object categorization.

An embodied account of abstract language might further argue that our understanding of abstract concepts like change in quantity is based on our experience with concrete, tangible domains like change in physical height, because the two are systematically correlated in experience (Grady, 1997; Lakoff & Johnson, 1980). Indeed, much of the time when we experience a change in quantity or compare or evaluate quantity of physical entities, physical height correlates with quantity. For example, when water is poured into a glass, the increase in the amount of water goes along with the increase in height of the waterline, and the same is true of masses and piles of things. Thus, our understanding of abstract notions like quantity could be inextricably linked to its perceptual or motor correlates. Perhaps, when we deal with abstract concepts like quantity, even when applied to non-physical entities, we still engage our perceptual systems in reflection of their tight coupling with abstract notions in experience. More specifically, perhaps change of quantity verbs activate visual up-down imagery in the same way literal change of height verbs do.

### 5.1. Method

Abstract verbs were selected from a single semantic field. All verbs expressed a change in quantity—either an increase, such as *increase* and *double*; or a decrease, such as *decrease* and *lessen*. They only encoded change in quantity (and could not independently denote change in height), using language primarily associated with quantity (i.e., non-metaphorical abstract motion). Sentences were constructed using these abstract verbs along with sentential subjects that denoted abstract quantifiable entities, drawn from the same group as those used with the metaphorical sentences in Experiment 3. This yielded sentences like those in the following:

- a. The figures doubled. [Abstract Up]
- b. The percentage decreased. [Abstract Down]

Because the abstract verbs used here do not denote any literal upward or downward motion, it is critical to determine that they are nevertheless strongly associated with the vertical axis. In the norming study, where participants were asked to rate verbs for upness or downness, they systematically assigned verbs denoting increases, like *increase* and *double* high Up ratings and verbs denoting decreases high Down ratings. Indeed, the verbs in the two abstract conditions were significantly different from each other in upness rating,  $F(1, 28) = 86.49, p < .001$ ; and downness rating,  $F(1, 27) = 149.78, p < .001$ . By contrast, the nouns in abstract up versus down sentences were not rated differently in upness:  $F(1, 28) = 0.03, p = .87$ ; or in downness:  $F(1, 27) = 0.07, p = .79$  (Table 1). Abstract sentences in the two conditions showed no significant difference in the RTs:  $F(1, 28) = 1.54, p = .23$ ; or in the meaningfulness ratings:  $F(1, 28) = 0.01, p = .94$ .

Table 8  
 Mean reaction time in milliseconds for object categorization in upper and lower quadrants of the screen for Up and Down sentences in Experiment 4, by sentence

Sentences	Object Up	Object Down
Abstract Down		
The Ratio Lessened.	593	507
The Quantity Dwindled.	549	505
The Indicators Weakened.	647	578
The Percentage Decreased.	583	700
The Value Diminished.	504	630
Abstract Up		
The Fees Expanded.	670	592
The Rating Improved.	642	595
The Price Redoubled.	637	589
The Figures Doubled.	540	556
The Numbers Increased.	515	640

699 The experiment was conducted using the same method as those described previously, and  
 700 was run together with Experiment 1.

701 *5.2. Results*

702 By contrast with the literal up and down sentences, the means for the abstract sentences  
 703 show no interference effect (Table 3). Indeed, a participant analysis of RTs following abstract  
 704 sentences showed no significant interaction of sentence direction with object location,  $F(1,$   
 705  $63) = 0.13, p = .72$ ; partial  $\eta^2 = 0.002$ . There were no significant main effects of sentence  
 706 direction or object location either. The individual items in the abstract condition (Table 8) did  
 707 not display the polarization seen in the responses to individual items in the literal sentences  
 708 in Experiments 1 and 2: as many abstract down sentences (3 out of 5) yield longer response  
 709 times whether the object is displayed in the upper or the lower part of the visual field.

710 *5.3. Discussion*

711 Despite being systematically associated with upness or downness, the abstract verbs used  
 712 in this experiment did not yield selective interference on the object categorization task. This  
 713 provides further evidence that the outcomes of the Experiment 1 and Experiment 2 did not  
 714 result simply from lexical attributes of the constituent words in the sentences—something like  
 715 a [+UP] or [+DOWN] feature. The abstract up verbs were strongly up-associated, and the ab-  
 716 stract down verbs were strongly down-associated, at least as measured by the norming data; yet  
 717 these aspects of their semantics were not sufficient for them to interfere with visual object cat-  
 718 egorization. There is a straightforward explanation for the presence of an interference effect in  
 719 the first two studies and its absence in the last two. Namely, the scenes described by the first two  
 720 involved actual events occurring in one location or the other, whereas those described by the

last two did not. It would thus seem to be the construction of a mental representation of the de- 721  
scribed scene, rather than purely lexical semantics, that drives the measured interference effect. 722

Given the finding in this fourth study, that abstract language about change in quantity 723  
does not trigger visual imagery as measured by interference on visual perception, we are 724  
left without an answer to the question of how abstract language is understood and, more 725  
generally, how abstract concepts are represented. Indeed, there is a great deal of variability 726  
in experimental results pertaining to the processing of abstract and metaphorical language. 727  
Although there are reliable spatial effects during abstract language processing in orientation 728  
judgment (Richardson, Spivey, & Cheung, 2001) and Perky-type tasks by axis (Richardson 729  
et al., 2003, Experiment 1), spatial effects are not observed in a Perky-type task by location 730  
(our Experiment 4) or in a picture recall task (Richardson et al., 2003, Experiment 2). 731

Despite this variability in experimental results, it has been widely suggested that we base 732  
abstract thought and language on concrete thought and language (Barsalou, 1999; Barsalou 733  
& Wiemer-Hastings, 2005; Lakoff, 1987). For instance, change in quantity is understood in 734  
terms of change in height. This study shows that it is not straightforwardly the case that a 735  
particular abstract domain is processed in exactly the same way as the concrete domain it 736  
is supposedly related to. Of course, this should not be particularly surprising. If individuals 737  
understanding abstract language enacted mental imagery that was not qualitatively different 738  
from imagery performed during literal language processing, this would be a confusing state of 739  
affairs for comprehenders indeed. Because we know that in understanding language, people 740  
are not prone to confusing changes in quantity of abstract numbers with change in height of 741  
physical objects, the processing of these different domains must differ in some ways. 742

It remains to be seen exactly what processes underlie abstract language understanding, but 743  
the absence of an interference effect observed here does not imply that the embodied account 744  
for abstract language understanding and abstract concept grounding is incorrect. There may be 745  
other factors that obscure a measurable interference effect with abstract sentences, entertained 746  
in section 6. A key finding of this final experiment, however, is that where Richardson et al.'s 747  
(2003) earlier work showed that abstract sentences yield interference effects on categorizing 748  
objects in the same axis, we found no effect of abstract sentences on categorizing objects 749  
in the same location. In addition, the results of Experiment 1 showed significant effects for 750  
literal concrete sentences; but, Richardson et al.'s concrete sentences appeared not to produce 751  
significant effects, albeit in statistically unlicensed post hoc tests. In the last study, we consider 752  
possible explanations for these divergences and test the idea that the differences lie in the detail 753  
of the mental imagery driven by concrete versus abstract language. 754

## **6. Experiment 5: abstract verbs and nouns** 755

Although the present work and Richardson et al.'s (2003) differed along several dimensions, 756  
the most obvious one is the assignment of sentences to different conditions. The original study 757  
took upward- and downward-directed sentences as belonging to the same condition (contrasted 758  
with horizontal sentences) and categorized all responses to objects appearing either in the upper 759  
or the lower part of the screen as belonging to the same condition (contrasted with right- or 760  
left-appearing objects). In other words, the sentence and image stimuli were specific to the 761

762 axis of concrete or abstract motion. By contrast, the current study pulled apart the up and  
763 down conditions in sentences and object responses. This offers a straightforward explanation  
764 for the difference in responses to literal sentences in the two experiments.

765 Given that we have seen in this work that literal up sentences interfere with visual processing  
766 in the upper part of the screen, and down sentences interfere with the lower part of the visual  
767 field (Experiments 1 and 2), it is not at all surprising that grouping all these responses  
768 together (as was done in Richardson et al., 2003) would eliminate any effects. After all, up  
769 sentences (possibly about one half of the sentences in the vertical condition) would result  
770 in slower responses to objects in the upper part of the screen (one half of the objects in  
771 that condition), whereas down sentences (the remaining sentences in that same condition)  
772 would interfere with the other half of the object stimuli—those in the lower position. The  
773 two effects could cancel each other out, resulting in no significant effect. By comparison,  
774 this study, which investigated not just axes but more particularly locations along those axes,  
775 did not see such effects obscured, and the results were thus clearly significant for concrete  
776 sentences.

777 By contrast, there are several candidate explanations for why abstract sentences showed  
778 a significant interference effect by axis in the original study (Richardson et al., 2003) but  
779 no location-specific interference in our Experiment 4. The most prominent one is based on  
780 this same structural difference between the experiments, placing up and down in different  
781 conditions or collapsing them into a single vertical axis condition. Perhaps, as Richardson  
782 et al. showed, abstract sentences do trigger mental imagery, but imagery that is not specific  
783 to particular locations so much as to axes—that is, abstract language imagery may be less  
784 spatially precise, while still retaining an imagistic component. This would explain why abstract  
785 language yields measurable interference effects when up and down are collapsed together  
786 and the entire vertical axis is treated as a condition. It would also explain why a study  
787 like Experiment 4 in which objects located in the upper and lower regions are placed in  
788 separate conditions would show no such interference because the abstract motion sentences  
789 are not incompatible with any of the presented objects, all of which appear in the vertical  
790 axis.

791 Some support for this account comes from evidence that axes and specific locations are rep-  
792 resented distinctly in the human cognitive system (Logan & Sadler, 1996). Carlson-Radvansky  
793 and Jiang (1998) have shown that individual words like *above* may activate an entire axis,  
794 presumably as contrasted with location-specific words like *up*. McCloskey and Rapp (2000)  
795 have similarly shown that axis and direction can dissociate in particular neurological disor-  
796 ders. A participant they studied had lost the ability to ballistically reach for targets (thus,  
797 had lost location specificity) but preserved the ability to interact with the correct axis along  
798 which the object was located. Similarly, Landau and Hoffman (2005) have shown that children  
799 with Williams Syndrome have difficulties with direction but not axis of orientation. Thus, it  
800 is reasonable to conclude that object location may be represented separately from axis of  
801 orientation, and as such the two different systems might be available to be recruited separately  
802 by concrete versus abstract language processing.

803 We tested this explanation using the same methodology as in Experiment 4, except that the  
804 critical abstract sentences were now followed by objects appearing not only in the upper and  
805 lower parts of the screen, but also on the right and left. This required us to double the number

of abstract up and down sentences using the same template as in Experiment 4. If we found 806  
an effect of axis but not quadrant—that is, if abstract sentences yielded slower response times 807  
to object categorization in the upper and lower parts of the screen than in the left and right 808  
parts—this would replicate Richardson et al.’s (2003) findings and support the hypothesis 809  
that abstract sentences are simulated with less detail than concrete ones. 810

6.1. Method 811

Although our main focus was on abstract sentence processing, we also included metaphori- 812  
cal and noun-based sentences as controls, along with filler items. Each participant saw each of 813  
the three types of sentences. The concrete verb-manipulated sentences were not included, as 814  
this would have led to excessive repetition of verbs in the verb-manipulated and metaphorical 815  
conditions. 816

The original sets of sentences used in the first four experiments included only five verbs for 817  
each condition, with each sentence that used these verbs repeated twice for each participant. 818  
In order to present targets in each of the four quadrants of the screen, we needed to increase 819  
our stimulus set. We increased the number of verbs in each condition from five to eight, 820  
selecting an additional three verbs (or nouns) from those having the highest ratings in upness 821  
or downness from the previous norming study described in Experiment 1. We then doubled 822  
the number of stimuli for each condition by using each verb twice but with a different noun 823  
for the metaphorical and abstract conditions, and each noun twice for the noun sentences, with 824  
a different verb. An example abstract sentence pair is shown in the following (9). The verb 825  
*failed* was rated as strongly downward associated. Unlike the previous studies, participants 826  
saw each sentence (e.g., 9a or 9b) only once. 827

- a. The argument failed. 828
- b. The policy failed. 829

Unbiased nouns for the metaphorical and abstract sentences, and unbiased verbs for the 830  
noun sentences, were chosen from the norms to have low ratings for up or downness. We also 831  
included a few words that were not in the original norms, in order to construct new intransitive 832  
sentences that made sense. When this was done, care was taken not to include words that had 833  
an intuitively obvious association with the vertical or horizontal axes. The list of the abstract 834  
sentences used in this experiment are included in the appendix. 835

The presentation of stimuli was globally the same as in Experiments 1 through 4. However, 836  
in those experiments, only filler sentences preceded visual targets appearing in the left or 837  
right regions of the screen, whereas in this experiment horizontal object presentation followed 838  
critical experimental sentences. This experiment used one list, with the pairing of sentence 839  
type to item target randomly assigned for each participant, but with each of the four possible 840  
target locations (up, down, left, or right) appearing with equal frequency for each sentence 841  
type within participants. 842

Responses were collected using an E-Prime button box instead of the keyboard used in 843  
Experiments 1 through 4. Sentences were recorded by a native speaker of British English. 844



845 6.2. Results

846 Fifty native speakers of English from the University of Sussex community took part, in  
 847 exchange for course credit in a research methods class. All participants had above 85%  
 848 accuracy in target discrimination and 88% accuracy in the questions testing comprehension.  
 849 Outlier removal was the same as in Experiment 1.

850 The RTs for the left and right target locations were collapsed together in the analysis as  
 851 the horizontal axis, and the up and down targets formed the vertical axis. If these abstract  
 852 sentences yield mental imagery along the entire vertical axis, we should see longer RTs to  
 853 categorize objects when they appear after such sentences in the vertical axis than the horizontal  
 854 axis. However, analysis of just the abstract sentences with a repeated-measures ANOVA by  
 855 participants showed no significant difference in responses to the horizontal and vertical targets,  
 856  $F(1, 48) = 0.61, p = .44$ ; partial  $\eta^2 = 0.013$ . There was also no effect of target object location  
 857 when the metaphorical and noun-manipulated sentences were included; a 2 (horizontal  
 858 or vertical dimension)  $\times$  3 (abstract, metaphorical, or noun sentences) repeated-measures  
 859 ANOVA showed no main effect for horizontal or vertical object locations,  $F(1, 48) =$   
 860  $1.11, p = .30$ ; partial  $\eta^2 = 0.023$ ; and no significant interaction between sentence type and  
 861 object axis,  $F(2, 48) = 0.05, p = .94$ ; partial  $\eta^2 = 0.001$ . As a confirmation of the results  
 862 of Experiment 4, there was no significant interaction between sentence direction and up or  
 863 down object location for the new set of abstract sentences:  $F(1, 48) = 0.23, p = .88$ ; partial  
 864  $\eta^2 = 0.0$ .

865 One discrepancy between the previous set of studies is that the RTs were globally quicker  
 866 than Experiments 1 through 4, with a mean response of 289 msec in Experiment 5, compared  
 867 with 546 msec in Experiments 1 through 4. The reasons for this difference remain unclear  
 868 to us. The experiment was run on a different computer to the other studies, using a button  
 869 box instead of a keyboard, and with a different population (British vs. Hawaiian university  
 870 students). It is assumed that a combination of factors led to the shorter RTs, as the only main  
 871 difference in design between the studies was the inclusion of more sentence types. Although  
 872 no significant effect of axis was found, it is noted that for all three types of sentences the  
 873 RTs were slower for the vertical targets than the horizontal targets (see Table 9), although this  
 874 difference was very small—between 3 to 5 msec, and the level of unsystematic variability  
 875 meant that differences of this size were not enough to be statistically significant.

Table 9  
 Mean RT in milliseconds for object categorization in upper and lower quadrants of the screen for noun, metaphorical, and abstract sentences in Experiment 5

Sentence	Object in Vertical Axis			Object in Horizontal Axis			Difference (msec)
	Mean RT	SD	SE	Mean RT	SD	SE	
Noun	286	95	14	289	80	11	3
Metaphorical	288	86	12	291	90	13	3
Abstract	287	94	13	292	78	11	5

Note.  $N = 50$ . RT = reaction time.

### 6.3. Discussion

876

The results of Experiment 5 showed there was no interference effect for abstract sentences 877 by axis. They also replicated the finding of Experiment 4, showing that abstract sentences 878 yield no interference effect by up versus down location. The hypothesis that the differences 879 between the results of Experiment 4 in this work and Richardson et al. (2003) were due to 880 differences in the detail of the imagery prompted by concrete and abstract language is not 881 supported. Thus, it remains to be determined what caused the discrepancy between Richardson 882 et al.'s work and our Experiments 4 and 5. 883

One possible explanation for the absence of an effect with abstract sentences in our Experi- 884 ments 4 and 5, but the presence of such an effect in Richardson et al.'s (2003) work, relies on 885 differences in the abstractness of the stimuli in the two studies. In Richardson et al.'s work, 886 abstract sentences included verbs rated as abstract in the MRC Psycholinguistic database. This 887 selection method may have inadvertently resulted in a small number of relatively concrete 888 verbs; perusing the verbs in their study yields several candidates like *argue*, *rush*, *give*, and 889 *rest*. These verbs were combined with arguments that were very concrete—sentential subjects 890 always denoted people like *the storeowner*, *the husband*, or *the jogger*. The combination of 891 even relatively abstract verbs—like *want*—with concrete arguments—like *the child* and *the 892 cake*—results in sentences that could easily yield mental imagery of concrete situations. In 893 this example, an imagined scenario in which a child wants cake might involve a child looking 894 covetously at some cake in a spatial arrangement that is probably horizontal. Because ab- 895 stract sentences in the original study contained linguistic elements that might have made the 896 scenes they described concretely imageable, those images might have been responsible for 897 the interference effect observed with these abstract sentences. 898

By contrast, abstract sentences in the current study (Experiments 4 and 5) were more 899 abstract. All verbs (Table 8 and the Appendix) denoted change in quantity (some, such as 900 *expand*, are inevitably somewhat concrete as in Richardson et al.'s, 2003, study). However, 901 the nouns in the sentences are all abstract and describe quantitative measures like *quantity*, 902 *ratio*, and *measures*. As a result, it is subjectively more difficult to imagine a concrete scene 903 in which the scenes these sentences describe would be grounded than it is for the abstract 904 sentences in the original study. This could be responsible for the difference in findings in 905 the two studies—perhaps abstract language only yields measurable imagery effects when it is 906 straightforwardly interpreted as referring to spatially concrete scenes. We leave this possibility 907 open for investigation in future work. 908

## 7. General discussion

909

Processing sentences denoting events that would tend to take place in a particular part of 910 a perceiver's visual field yields interference on actually using the same part of the real visual 911 field, as measured by decreased performance in an object categorization task. This is true 912 whether the location of the event is denoted by a verb of motion (Experiment 1) or supplied 913 by connotational semantics of a sentential subject (Experiment 2). However, having an up- 914 or down-associated lexical item in a sentence does not suffice to produce interference. The 915

916 sentence must encode a scene literally involving the relevant location in the visual field, as  
917 metaphorical uses of motion verbs (Experiment 3) and abstract verbs that are nonetheless  
918 associated with upness or downness (Experiments 4 and 5) yield no significant interference  
919 effect, either at a specific level of detail (up or down; Experiment 4) or at a more general  
920 level of detail (vertical or horizontal axis; Experiment 5). We can conclude from this that it is  
921 not lexical priming that yields the interference but rather the performance of mental imagery  
922 corresponding to the meaning of an utterance.

923 One specific point about these experiments and the comparisons with previous work is  
924 worth taking up before we move on to a more general discussion of the place of imagery in  
925 language use. This is the question of why sentences in the first experiment, which denoted  
926 motion in a direction, interfered with static images of objects in particular locations. We  
927 used static visual stimuli for two reasons. The first was to enable comparisons with the work  
928 by Richardson et al. (2003), more of which follows below. The second was that we were  
929 concerned that moving objects would make it easier for participants to discern the relationship  
930 between the sentences and the visual perception task. The fact that we found significant effects  
931 despite this difference between the motion described by the sentences and the lack of motion  
932 in the visual stimuli suggests that the mere use of a particular location in the visual field can  
933 produce interference.

934 The findings reported in the foregoing studies provide new evidence suggesting that under-  
935 standing spatial language leads individuals to activate internal simulations of the described  
936 scenes. Although the selective interference of language processing on visual perception does  
937 not imply that such mental simulation is required for language understanding, it does imply  
938 that it is unconscious and automatic. Various authors have suggested different roles for the  
939 construction of a mental simulation on the basis of language, using detailed modal knowledge.  
940 One critical role of imagery is to produce detailed inferences (Narayanan, 1997), which can  
941 both allow an individual to gain a rich notion of the utterance's content, such as a situation  
942 model of the described scene (Zwaan, 1999), as well as to prepare the individual to understand  
943 future utterances or to respond relevantly. The construction of a mental simulation might  
944 also prepare the individual for situated action (Bailey, 1997; Barsalou, 1999; Glenberg &  
945 Kaschak, 2002). Finally, some language may be disambiguated only through the performance  
946 of imagery (Bergen & Chang, 2005).

947 Various theories of language rely heavily on perceptually and motorically grounded repre-  
948 sentations as the backbone for the language understanding process. Of particular note, Kaschak  
949 and Glenberg (2002) argued that language understanding proceeds through the meshing of **Q12**  
950 simulation constraints from language, and the subsequent mental simulation of afforded ac-  
951 tions, to prepare for situated responses. Zwaan (1999, 2004) argued similarly that language **Q13**  
952 comprehension proceeds through the construction of modal mental models, and Barsalou  
953 (1999) suggested that language hooks into simulators—systematic patterns of reactivation of  
954 representations of perceptual and motor experiences. What all these approaches share is a  
955 recognition of the importance of mental simulation in the process of language understanding.  
956 However, none of them are actual theories of how the individual linguistic items that make  
957 up an utterance directly produce a mental simulation, especially given the complexities of  
958 linguistic structure, although Kaschak and Glenberg made some progress with regard to how  
959 grammatical constructions contribute to mental simulation.

Up to the present, one of the main gaps in theories of language understanding based on mental simulation is explaining the precise ways in which language triggers simulation and what aspects of simulation it triggers. Kaschak and Glenberg (2002 for example, view the construction of an embodied simulation as arising from the *meshing* of simulation constraints imposed by pieces of language, but very little is known about how exactly this might take place or what aspects of simulation can be triggered by what sorts of language. Cognitive linguists have documented a broad range of possible functions of grammatical and lexical items. For example, it appears that various sorts of language, from modal verbs like *make* and *let* to prepositions like *despite* and *from*, are intuitively associated with simple notions of the application or non-application of force (Talmy, 2000). A function of various grammatical structures, like subjects and topic markers, appears to be to raise certain elements to prominence as the foreground by contrast with others that remain in the background (Lakoff, 1987; Langacker, 1987; Talmy, 2000). Although cognitive linguistic work is based largely on introspection and text analysis, it provides many useful insights into language use and representation and serves as an extremely rich source for empirically testable potential functions of linguistic items.

Q14

Work like the experiments described here can begin to tell us a little bit more about exactly how language drives simulation. One thread of work attempting to wed the observation that simulation is a central element in language understanding with the details of how specific linguistic elements drive simulation, as inspired by the work in cognitive linguistics described above, is “embodied construction grammar” (Bergen & Chang, 2005; Bergen et al., 2004; Feldman, 2006). The basic idea of embodied construction grammar, a computational model of language understanding, is that linguistic elements (from lexical items to grammatical markers to phrasal patterns) are pairings of some linguistic form with specifications for mental simulations to be performed when they are used. In the simplest cases, words that denote actions or perceivable entities drive the simulation to enact imagery of those actions or entities. Similarly, grammatical constructions place constraints on the simulation—indicating what type of event should be simulated, from what perspective, or with what in the foreground. As in Glenberg’s model, the simulation constraints of the various linguistic constraints must be meshed or *bound* together to produce a coherent simulation for an utterance. We anticipate that future work will further elucidate the contributions that individual words, as well as grammatical structures, make to the construction of mental imagery during language understanding.

Q15

Visual interference effects produced by linguistic input are reliable and replicable in a number of methodological permutations. These findings as a whole provide evidence that perceptual systems—in particular the visual system—are unconsciously and automatically engaged in the process of natural language understanding. Given that spatial imagery is automatically engaged during language use, it seems that a complete account of how words and utterances are understood requires knowing how they drive imagery. The same may hold of grammatical markers and sentence patterns (Bergen & Chang, 2005; Glenberg & Kaschak, 2002). More broadly, the observation of language driving imagery suggests yet another way that embodied human experience shapes language processing. Our similar bodies and experiences yield shared imagery, a common currency that facilitates effective communication

1004 **Notes**

- 1005 1. The relatively small number of sentences of each type could, in principle, be remedied  
1006 by using the words *up* and *down* in sentences. We chose to avoid these words for several  
1007 reasons. First was the possibility that participants would recognize these recurring words  
1008 in the experiment and guess its purpose. We were also concerned with potential direct  
1009 effects of the words *up* and *down* on participants' responses. For example, seeing those  
1010 words might result in participants orienting overt attention to that part of the visual field,  
1011 which would counteract the expected effect. Moreover, if included, *up* or *down* could  
1012 themselves be argued to be responsible for any observed effects rather than the interpre-  
1013 tation of the sentence as a whole (which we tested by contrasting Experiments 1 and 3).  
1014 2. Replacing outliers with values at a set distance from the subject's mean is also known as  
1015 "winsorizing" (Barnett & Lewis, 1978) and is commonly used in sentence processing  
1016 research. Although it may increase power in a small set of restricted cases, it globally  
1017 does not affect results of statistical analyses (Ratcliff, 1993). We chose to winsorize,  
1018 rather than eliminate outliers, due to the small number of items in each condition.

1019 **Acknowledgments**

1020 Our thanks to Larry Barsalou, Susanne Gahl, Art Glenberg, Terry Regier, Daniel  
1021 Richardson, Amy Schafer, and Nathaniel Smith for useful discussion and comments, as well  
1022 as to Steven Sloman, Arthur Markman, and three anonymous reviewers for careful and useful  
1023 commentary. Any errors or omissions are our own.

1024 **References**

- 1025 Bailey, D. (1997). *When push comes to shove: A computational model of the role of motor control in the acquisition*  
1026 *of action verbs*. Unpublished doctoral thesis, University of California, Berkeley. **Q16**
- 1027 Barnett, V., & Lewis, T. (1978.) *Outliers in statistical data*. New York: Wiley.
- 1028 Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–609.
- 1029 Barsalou, L. W., Simmons, W. K., Barbey, A., & Wilson, C. D. (2003). Grounding conceptual knowledge in  
1030 modality-specific systems. *Trends in Cognitive Sciences*, 7, 84–91.
- 1031 Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. Zwaan (Eds.),  
1032 *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). New  
1033 York: Cambridge University Press.
- 1034 Bergen, B. (in press). Experimental methods for simulation semantics. In M. Gonzalez-Marquez, I. Mittelberg, S.  
1035 Coulson, & M. J. Spivey (Eds.), *Methods in cognitive linguistics*. Ithaca, NY: Benjamins. **Q17**
- 1036 Bergen, B., & Chang, N. (2005). Embodied construction grammar in simulation-based language understanding. In  
1037 J.-O. Östman & M. Fried (Eds.), *Construction grammars: Cognitive grounding and theoretical extensions* (pp.  
1038 147–190). Amsterdam: Benjamins. **Q18**
- 1039 Bergen, B., Chang, N., & Narayan, S. (2004). Simulated action in an embodied construction grammar. In *Proceed-*  
1040 *ings of the 26th annual conference of the Cognitive Science Society*. **Q19**
- 1041 Bergen, B., Narayan, S., & Feldman, J. (2003). Embodied verbal semantics: Evidence from an image-verb matching  
1042 task. In *Proceedings of the 25th annual conference of the Cognitive Science Society*. **Q20**
- 1043 Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, 75, 1–28.
- 1044 Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, 13,  
1045 185–188.

- Bower, G., & Morrow, D. (1990). Mental models in narrative comprehension. *Science*, 247, 44–48 1046
- Carlson-Radvansky, L. A., & Jiang, Y. (1998). Inhibition accompanies reference frame selection. *Psychological Science*, 9, 386–391. 1047
- Craver-Lemley, C. E., & Arterberry, M. E. (2001). Imagery-induced interference for a visual detection task. *Spatial Vision*, 14, 101–119. 1049
- Ehrsson, H. H., Geyer, S., & Naito, E. (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part specific motor representations. *Journal of Neurophysiology*, 90, 3304–3316. 1051
- Feldman, J. (2006). *From molecule to metaphor: A neural theory of language*. Cambridge, MA: MIT Press. 1053
- Gibbs, R., Bogdonovich, J., Sykes, J., & Barr, D. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language*, 37, 141–154. 1054
- Glenberg, A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558–565. 1056
- Glenberg, A., & Robertson, D. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43, 379–401. 1058
- Grady, J. (1997). *Foundations of meaning: Primary metaphors and primary scenes*. Unpublished doctoral dissertation, University of California, Berkeley. 1059
- Johansson, R., Holsanova, J., & Holmqvist, K. (2005). What do eye movements reveal about mental imagery? Evidence from visual and verbal elicitations. In *Proceedings of the 27th Cognitive Science Conference*, Stresa, Italy. 1061
- Q21 Kaschak, M., & Glenberg, A. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language*, 43, 508–529. 1064
- Kaschak, M., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., et al. (2005). Perception of motion affects language processing. *Cognition*, 94(3), B79–B89. 1066
- Kersten, A. (1998). A division of labor between nouns and verbs in the representation of motion *Journal of Experimental Psychology: General*, 127, 34–54. 1068
- Kosslyn, S. (1980). *Image and mind*. Cambridge, MA: Harvard University Press. 1070
- Kosslyn, S., Ganis, G., & Thompson, W. (2001). Neural foundations of imagery. *Nature Reviews. Neuroscience*, 2, 635–642. 1071
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press. 1073
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press. 1074
- Landau, B., & Hoffman, J. E. (2005). Parallels between spatial cognition and spatial language: Evidence from Williams Syndrome. *Journal of Memory and Language*, 53, 163–185. 1075
- Langston, W. (2002). Violating orientational metaphors slows reading. *Discourse Processes*, 34, 281–310. 1077
- Lindsay, S. (2003). *Visual priming of language comprehension*. Unpublished master's thesis, University of Sussex. 1078
- Logan, G. D., & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. Peterson, M. Garrett, & L. Nadel (Eds.), *Language and space*. Cambridge, MA: MIT Press. 1080
- Q22 Matlock, T. (2004a). The conceptual motivation of fictive motion. In G. Radden & R. Dirven (Eds.), *Motivation in grammar* (pp. 221–248). Amsterdam: Benjamins. 1081
- Matlock, T. (2004b). Fictive motion as cognitive simulation. *Memory & Cognition*, 32, 1389–1400. 1083
- Q23 Matlock, T., & Richardson, D. C. (2004). Do eye movements go with fictive motion? In *Proceedings of the 26th annual meeting of the Cognitive Science Society*. 1084
- McCloskey, M., & Rapp, B. (2000). Attention-referenced visual representations: Evidence from impaired visual localization. *Journal of Experimental Psychology*, 26, 917–933. 1086
- Narayanan, S. (1997). *KARMA: Knowledge-based action representations for metaphor and aspect*. Unpublished doctoral thesis, University of California, Berkeley. 1088
- Narayan, S., Bergen, B., & Weinberg, Z. (2004). Embodied verbal semantics: Evidence from a lexical matching task. *Proceedings of the 30th annual meeting of the Berkeley Linguistics Society*, CA. 1090
- Q24 Nyberg, L., Petersson, K.-M., Nilsson, L.-G., Sandblom, J., Aberg, C., & Ingvar, M. (2001). Reactivation of motor brain areas during explicit memory for actions. *NeuroImage*, 14, 521–528. 1092
- Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology*, 21, 422–452. 1094
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510–532. 1095
- Q25 Richardson, D. C., & Matlock, T. (in press). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition*. 1096

- 1098 Richardson, D. C., Spivey, M. J., & Cheung, J. (2001). Motor representations in memory and mental models:  
1099 Embodiment in cognition. In *Proceedings of the 23rd annual meeting of the Cognitive Science Society* (pp.  
1100 867–872). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- 1101 Richardson, D. C., Spivey, M. J., McRae, K., & Barsalou, L. W. (2003). Spatial representations activated during  
1102 real-time comprehension of verbs. *Cognitive Science*, 27, 767–780.
- 1103 Spivey, M., & Geng, J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to  
1104 absent objects. *Psychological Research*, 65, 235–241.
- 1105 Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture  
1106 recognition. *Psychological Science*, 12, 153–156.
- 1107 Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge, MA: MIT Press.
- 1108 Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in  
1109 language comprehension. *Cognitive Science*, 28, 611–619.
- 1110 Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Do language comprehenders routinely represent the shapes  
1111 of objects? *Psychological Science*, 13, 168–171.

Appendix

Table of abstract sentences used in Experiment 5

---

Abstract Down Sentences

The Indicators Weakened.  
The Prospects Weakened.  
The Value Diminished.  
The Faith Diminished.  
The Quantity Dwindled.  
The Interest Dwindled.  
The Ratio Lessened.  
The Indicators Lessened.  
The Enthusiasm Decreased.  
The Demand Decreased.  
The Argument Failed.  
The Policy Failed.  
The Crowd Saddened.  
The Nation Saddened.  
The Agreement Broke.  
The Pact Broke.

Abstract Up Sentences

The Ratings Improved.  
The Market Improved.  
The Fees Doubled.  
The Inflation Doubled.  
The Price Redoubled.  
The Payments Redoubled.  
The Amount Multiplied.  
The Price Multiplied.  
The Figures Expanded.  
The Program Expanded.  
The Numbers Increased.  
The Ranking Increased.  
The Coalition Conquered.  
The Army Conquered.  
The Prosecution Won.  
The Law Won.

---