Spatial and linguistic aspects of visual imagery in sentence comprehension

Ben Bergen

University of Hawai`i at Manoa

Shane Lindsay

University of Sussex

Teenie Matlock

University of California, Merced

Srini Narayanan

International Computer Science Institute

Contact:

Ben Bergen

bergen@hawaii.edu

Department of Linguistics

569 Moore Hall

1890 East-West Rd.

Honolulu, HI 96822

http://www2.hawaii.edu/~bergen

1. Introduction

"thought is impossible without an image."

-Aristotle, On Memory and Recollection

Until the late 1950's, mental imagery was believed to occupy a special role in human thought. In the second half of the 20th century, however, imagery was back-grounded by approaches that favored objectivism and symbol manipulation. Over the course of the past two decades, imagery has 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, 2001) and categorization (Barsalou, 1999a). 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 recently converged on the notion that language understanding critically engages mental imagery, especially mental simulation. Cognitive linguistics, for instance, 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). Cognitive psychology has similarly discussed the importance of low-level perceptual and motor processes in language and other high level phenomena (Barsalou, 1999a; 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, has the potential to interfere with

(Richardson et al., 2003; Kaschak et al., 2005) or facilitate (Glenberg & Kaschak, 2002; Zwaan et al., 2002) the actual performance of actions or the perception of objects, depending on the task.

This paper, which focuses on the role of visual imagery in language understanding, 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, it 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). These results 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.

1.1. Mental simulation in language understanding

To demonstrate the influence of language on imagery, it is useful to consider the subjective experience of processing language associated with perceptual content. Answering questions like the following, for instance, may require mental imagery: What shape are a poodle's ears? What color is

the cover of *Cognitive Science*? Which is taller: a basketball hoop or a bus? (See also examples in Kosslyn, 1980). Critically, in answering such questions, people report that they are mentally picturing, and "looking at" named objects; that they are mentally rotating or otherwise manipulating these objects; that they are able to zoom in or out; and that they are combining imagined objects in a single visual picture (Kosslyn et al., 2001). And all of this visual imagery is triggered by verbal input.

Imagery, then, can be defined as experience resembling perceptual or motor experience, occurring in the absence of the relevant external stimuli in the case of perceptual experience, or without actual execution of motor actions in the case of motor imagery. Imagery has played a critical role in most theories of mind, starting at least as early as Aristotle. Modern investigations of imagery have demonstrated that it is integral to conceptual knowledge (Barsalou et al., 2003) and recall (Nyberg et al., 2001); can work unconsciously (Barsalou, 1999b); can be used productively to form new configurations (Barsalou & Prinz, 1997); and works by activating neural structures overlapping with (or a subset of) those used for perception and action (Ehrsson et al., 2003; Kosslyn et al., 2001).

Imagery has been argued in the literature on embodied cognition and especially cognitive linguistics to be critical to language. The shared central idea is that processing language activates internal representations of previously experienced events, or schematic abstractions over these (Lakoff, 1987; Langacker, 1991; Talmy, 2000). It is thus the (re)activation of modal (e.g., perceptual or motor) content associated with particular described scenes that serves as the "engine" of meaning. This mental simulation process has been argued to be useful in the production of detailed inferences on the basis of language input (Narayan, 1997); to prepare the understander for situated action (Bailey, 1997; Barsalou 1999b; Glenberg & Kaschak, 2002); to build a situation model of the described scene (Zwaan, 1999); and to allow disambiguation (Bergen & Chang, 2005). Globally, then, embodied approaches to language predict that understanding verbal input about

events that can be perceived or performed will result in an individual's tacit and automatic mental enactment of corresponding motor or perceptual imagery.

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

In the remainder of this paper, we concentrate on visual imagery evoked in response 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, since, 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 since language that describes upwards or downwards motion or location occurs pervasively within languages. Since 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 literally describes 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.

(1)   a.   The ground/roof shook.

        b.   The ant climbed/dropped.

c.   Stock prices <u>climbed</u>/<u>dropped</u>.

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

1.2. Linguistic Perky effects

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 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 or not, 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 2001). If this is the case, then interference of visual imagery on visual processing can be naturally used to investigate whether language processing also drives imagery. Rather than asking participants to imagine visual objects, experimenters can ask participants to process language hypothesized to evoke visual imagery of a particular type – of particular objects with particular properties, or of objects in particular locations. If visual language selectively activates visual imagery, then we should expect a Perky-type effect that results in interference of the visual properties implied by the language on processing of the displayed visual image.

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

Specifically, Richardson et al. (2003) suggested that processing language about concrete or abstract motion along different axes in the visual field (vertical versus horizontal) leads language understanders to conscript the parts of their visual system that are normally used to perceive trajectories with those same orientations. For example, a sentence like (2a) implies horizontal motion, while (2b) implies vertical motion. If understanders selectively perform vertical or horizontal visual imagery in processing these sentences, then when they are asked immediately after presentation of the sentence to visually perceive an object that appears in their actual visual field, they should take longer to do so when it appears on the same axis as the motion implied by the

sentence. Thus after (2a), a horizontal-motion sentence, participants should take longer to categorize an object as a circle or a square when it appears to the right or left of the middle of the screen (on the horizontal axis) than it should take them to categorize an object when it appears above or below the middle of the screen (on the vertical axis).

(2)   a.   The miner pushes the cart.

        b.   The ship sinks in the ocean.

An additional point of interest here concerns the nature of the sentences used. The experimenters were interested in the spatial orientation not just of concrete verbs, like *push* and *sink*, but also abstract verbs, like *respect* and *tempt*. They wanted to determine whether abstract events, just like concrete events, were selectively associated with particular spatial orientations. How abstract concepts are represented and understood is a critical question for all theories of meaning and understanding, but is particularly critical to simulation-based models, which rely on perceptual and motor knowledge. There are insightful discussions of how abstract concepts can be grounded in embodied systems elsewhere (Lakoff, 1987; Barsalou, 1999; and Glenberg & Robertson 2000, Barsalou & Wiemer-Hastings, 2005), and the topic is 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) that were presented on the screen in locations that overlapped with the sentences' implied orientation. After seeing a fixation cross for 1 second, participants heard a sentence, then, after a brief pause (randomly selected for each trial from among 50, 100, 150, or 200ms), 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 (one button each for 'circle' and 'square') as quickly as possible.

(3)   a.   [Concrete Horizontal] *The miner pushes the cart.*

        b.   [Concrete Vertical] *The plane bombs the city.*

        c.   [Abstract Horizontal] *The husband argues with the wife.*

        d.   [Abstract Vertical] *The storeowner increases the price.*

The results were indicative of a clear interference effect -- participants took longer to categorize objects on the vertical axis after vertical sentences versus horizontal sentences, and vice versa for objects on the horizontal axis. Intriguingly, post hoc tests (which the authors indicate are, strictly speaking, statistically unwarranted) showed that this interference effect was significant for abstract sentences, but not for the concrete sentences (see Section 6 and in particular Footnote 3 for details).

It is important to underline at this point that the expected (and observed) effect involved 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 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 (Stanfield & Zwaan, 2001; Zwaan et al., 2002; and Glenberg & Kaschak, 2002), facilitation is observed (Kaschak et al., 2005, and others cited in Bergen, To Appear).

Though Richardson et al. (2003) provided further insights into the relationship between imagery and language, it left several questions unanswered, questions that we will explore in the current paper. First, why would abstract sentences but not literal sentences generate the expected Perky effect? No simulation-based account of language understanding, nor any account of language understanding that we are aware of would predict such an asymmetry – that abstract but not literal spatial language should yield perceptual imagery.

Second, this study was not designed to tell us what linguistic elements in the sentences were yielding the observed effects. The sentences used different argument structures, including both transitive and intransitive structures, and had subjects and objects whose own vertical or horizontal associations were not controlled for.

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

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

(4)   a.   The ship sinks in the ocean.
        b.   The strongman lifts the barbell.

Thus the current state of affairs still leaves open the three questions identified earlier. Namely, (1) what linguistic cues trigger mental simulation, (2) what sorts of language (literal, metaphorical, abstract) result in mental simulation, and (3) how detailed is the mental simulation?

2. Experiment 1: Literal up and down motion sentences

Does language denoting literal motion in a particular direction drive visual imagery localized to the same part of the visual field? Our first experiment followed Richardson et al. (2003), but aimed to answer the outstanding questions of what linguistic elements drive simulation and how detailed it is. The design here controlled for the linguistic components of sentences, and separated the vertical

axis into distinct up and down regions. Based on prior work showing that the Perky effect is location-specific (Craver-Lemley and Arterberry, 2001), we expect that people will take longer to identify objects in the upper or lower part of the visual field following sentences denoting scenes that canonically take place in the same locations.

To reduce the range of possible linguistic factors influencing imagery, we used only intransitive sentences (sentences with only a subject noun phrase and a main verb) . All subject nouns in the critical trial sentences were normed for their up/down-relatedness. In addition, verbs were only selected from among those that, as determined by a norming task, denoted literal motion in a particular direction. This meant that only upwards and downwards motion could be used, as there do not exist verbs in English that denote rightwards or leftwards motion. Critical sentences thus fell into two directional conditions (up and down). Examples appear in (5) below.

(5)   a.   [Literal Up] *The mule climbed*

      b.   [Literal Down]: *The chair toppled*

2.1. Method

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

Participants wore headphones and sat in front of a computer screen. They heard sentences and looked at geometric shapes that were presented in one of four locations on the screen. They were instructed to quickly press one of two buttons to identify whether the shape was a square (press 'x') or a circle (press 'z'). Each trial began with a fixation cross that appeared in the middle of the screen for 1000msec. Next a sentence was played, followed by an ISI of 200msec (during which time the screen was blank). Then a circle or a square appeared in the top, bottom, left or right part

of the screen for 200msec. All objects appeared the same distance from the fixation cross at the center of the screen, along a central axis (e.g., objects in the upper part appeared directly over the fixation cross).

There were two types of trials: critical and filler. In critical trials, sentences were played denoting either upward motion or downward motion (5), and the object appeared in the upper or lower region. Filler trials were randomly interspersed to ensure that participants attended to the meaning of the sentences. Filler trials included a short yes/no comprehension question. For instance, the filler sentence *The branch split* was followed by the question *Did the branch break?* Filler trials included as many up- and down-related sentences as appeared in the critical trials, but all of these were followed by an object on the left or right – all of these sentences were selected from among the sentences discarded through the norming study.

The constraints imposed by this design, that only intransitive verbs denoting upward or downward motion could be used, translated into a relatively small number of candidate verbs. In English, there are only 5 to 10 verbs denoting either upward or downward motion, each. Because of the small number of possible verbs of each type, the entire list of sentences was presented twice to each participant, making the object location (up or down) a within-subjects factor. To ensure that there was distance between the two instantiations of each critical sentence, the experiment was broken into two halves, each of which contained all critical sentences, in a random order. The order of the two halves was manipulated to create two lists. Participants were randomly assigned to one of these lists.

2.2. Norming

In constructing stimuli, we conducted a norming study to ensure that the critical sentences had several properties. For each type of sentence, we aimed to include sentences in the up condition

that were no more or less meaningful than sentences in the down condition, and to have as little difference as possible in processing time between the two groups of sentences. Second, and more critically, we wanted to ensure that the sentences, which had only a subject and a verb, differed in terms of their upness or downness only because of one manipulated word. So the sentential subjects used in the literal sentences in this experiment were constrained to be equally neutral for their up-down association (e.g., *chair* and *donkey*), while the verbs denoted significantly different up/down meanings (e.g., *climb* and *descend*).

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

The critical stimuli in the upness or downness rating task included verbs that the experimenters hypothesized to denote motion events canonically going upwards or downwards, and nouns denoting objects canonically located above or below an observer's head, and the sentences in the reading and meaningfulness part of the norming study were constructed from these words. In addition, each group of participants saw half of the proposed filler sentences, which were expected to be meaningful, and the other half with the verbs and participant nouns randomized across sentences, which were thus unlikely to be meaningful. Finally, each participant saw 15 sentences with transitive verbs used intransitively, which were also unlikely to be judged meaningful.

One participant was removed from the norming study analysis for having a mean RT more than 2 s.d. greater than the grand mean. We also removed all trials with reaction times less than 350ms, as these sentences were unlikely to have been thoroughly understood.

In selecting sentences for the main experiment, we eliminated all sentences with extremely fast or slow reaction times, low meaningfulness ratings, nouns with strong up or down associations, or verbs without strong up or down associations. This left five sentences in each critical condition.[1] The mean up-ness and down-ness ratings for the nouns selected for the main study are shown in Table 1. The nouns in literal up sentences were not significantly more up-related than those in literal down sentences: $F(1,28)=0.545$; $p=0.466$. Nor were they significantly more down-related (though the effect here approached significance): $F(1,27)=3.559$; $p=0.07$. Turning to the verbs, it was crucial that the verbs used in two conditions differed from each other in terms of their up/down-ness. Overall, verbs were classified as expected: The verbs in the two literal conditions differed significantly in their upness ratings ($F(1,28)=117.649$; $p<0.001$) and their downness ratings ($F(1,27)=134.537$; $p<0.001$).

Also of interest are the mean reading times and meaningfulness ratings, shown in Table 2. Repeated measures ANOVAs revealed a reliable difference in reading times, $F(1,28)=12.39$, $p<0.01$, and a marginally significant difference in meaningfulness , $F(1,28)=4.095$; $p=0.053$. While it is certainly not desirable to have such differences between conditions, it was a necessary artifact of the design, as very few verbs exist in English that can denote intransitive upwards motion. This will be discussed in more detail in the Results section.

2.3. Results

Only participants who answered the sentence comprehension questions with at least 85% accuracy were included in analysis - this eliminated one participant. Another participant was excluded for answering the object categorization questions with only 79% accuracy. None of the remaining

participants performed at less than 90% accuracy on the critical trials. Responses that were 3 s.d. above or below the mean for each participant were removed and replaced with values 3 s.d. above or below the mean for that participant.[2] This resulted in changes to less than 1% of the data.

The mean reaction times for the literal sentences displayed in the first two data rows of Table 3 show a clear interaction effect of the predicted kind. Objects in the upper part of the visual field are categorized faster following literal down sentences than they are following literal up sentences, and the reverse is true for visual objects in the lower part of the visual field (though this latter effect does not appear to be as strong). Repeated measures subject ANOVA showed the interference effect for literal sentences through a significant interaction between sentence direction (up or down) and object location (up or down) (F (1,63) = 5.028; p < 0.05). There were no significant main effects of sentence type or object location. With only five items in each condition, it would be unrealistic to expect an ANOVA using items as a random factor to show significance. Moreover, since the set of stimuli in each condition effectively constitutes the population of relevant items, and are not random samples from that population, it would not make sense to perform such an analysis in any case. As shown in Table 4), however, all up sentences had longer RTs in the Up Object Condition than in the Down Object condition (by at least 30ms), suggesting that the interference effect holds for all the Literal Up sentences. Similarly indicative of interference, three out of five of the Literal Down sentences had longer RTs in the Down than in the Up condition. Looking at the items individually, it seems that the interference effect stronger with Literal Up sentences, which yielded much slower response times on average on objects in the upper position than those in the lower position.To deal with the problem of a small set of potential verbs, the design of the current study presented each critical sentence once with the visual stimulus in the upper region and once with the visual stimulus in the lower region. Because the repetition of stimuli runs the risk of inducing carryover effects (e.g., participants develop different strategies for responding to stimuli they have or have not seen already), we performed a post hoc analysis to

determine whether such effects accounted for the results reported here. To do this, we analyzed the data from the first half of the experiment, which included only the first presentation of each sentence. The results, seen in Table 5, are not statistically significant (F (1,63), $p < 1$), as might be expected, given the low number of stimuli per condition per participant (2.5). However, the trend is in same direction as the full results, suggesting that carryover effects were not responsible for the critical Perky-like interaction we observed.

2.4. Discussion

The significant interaction effect observed here with literal sentences of upwards or downwards motion leads to two important conclusions. The first involves the specificity of the imagery associated with these sentences. While it is known (Richardson et al., 2003) that the axis of motion of a sentence is accessed during language processing, the current study provides evidence that the spatial grain of visual imagery is in fact even more detailed than this. Since sentences denoting upwards and downwards motion selectively interfered with categorizing objects in the same part of the visual field, we can see that motion imagery in response to these sentences is specific to the location in which the content of the utterance would take place, not just the axis.

Second, unlike the post hoc report on Richardson et al.'s (2003) results, we observed a reliable interaction with concrete sentences denoting physical motion. This finding is more squarely in line with what is predicted by theories of perceptual simulation in language understanding - that literal language about space should be processed using those neuro-cognitive systems responsible for perceiving the same aspects of space. As suggested in the introduction, these results suggest that the lack of an effect for concrete sentences in Richardson et al. (2003) may have resulted from the conflation of the up and down directions into a single level. As we have seen here, sentences denoting upwards motion result in interference in the upper part of the visual field. Thus, it is not

17

surprising that when upwards- and downwards-oriented sentences are combined in a single condition, their effects might cancel each other out.

The effect we observed here was especially strong for sentences denoting upwards motion. Why might up sentences and not down sentences show this effect? One plausible explanation is that the difference results from the slightly (though not significantly) greater time it took participants to process the upwards motion sentences. Perhaps they had not completed the comprehension process at the point when the visual object was presented – in this case, continued sentences imagery would yield a greater interference effect.

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

Regardless of the details of this effect, the crucial manipulation that yielded it was the use of verbs that were strongly associated with upward or downward motion. From the simulation-based perspective, this is not particularly surprising, since verbs of motion are supposed to indicate processes and relations holding of referents. What would happen, though, if nouns were manipulated while verbs were held constant? Do nouns denoting objects that are canonically associated with the upper or lower part of the visual field yield the same sort of interference? This is the topic of the next study.

3. Experiment 2: Up- or down-associated subject nouns

In Experiment 1, we found a significant interference effect when a sentence's verb denoted motion 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 those in (6)) affected how long it took participants to perform the object 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 an egg in a pan. These results imply that shape and orientation of objects are represented in mental imagery during language understanding.

(6)   a.   *The man hammered the nail into the floor*

        b.   *The man hammered the nail into the wall*

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 et al, 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 (above or below an observer, for example) 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

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), for example, *The cellar flooded* and *The ceiling cracked*. The sentences were constructed from items selected from the norming study described in Experiment 1. In the norming study, the Up and Down sentences showed no significant difference in reaction time: $F(1,27)=0.889$; $p=0.354$ or in meaningfulness: $F(1,27)=2.602$; $p=0.118$ (see Table 2).

Moreover, the verbs in the two noun conditions did not differ significantly in either their upness ratings ($F(1,28)=0.131$; $p=0.720$) or their downness ratings ($F(1,27)=0.009$; $p=0.926$) (see Table 1). By contrast, the nouns in the up versus down sentences were highly differentiated in terms of upness: $F(1,28)=215.158$; $p<0.001$ and down-ness: $F(1,27)=132.307$; $p<0.001$. These results serve to ensure that any interference effects observed on the object categorization task

20

would result from the differences in the up or down associations of nouns alone, not in differences between the verbs.

3.2. Results

Response times from two participants whose mean response times fell 2 s.d. above the mean for all participants were removed. In addition, response times for two other participants were removed for answering the comprehension questions (to filler questions) with less than 80% accuracy. In the remaining data set, responses more than 3 s.d. from each participant's mean RT were replaced with values 3 s.d. from their mean. This resulted in the modification of less than 1% of the data.

Considering only correct responses, the means were as shown in Table 3. As with the verb manipulation in Experiment 1, there was interference in the predicted direction between sentence direction and object location. Indeed, a repeated-measures by-participants ANOVA showed a significant interaction between object location and sentence direction ($F(1,58)=5.756$; $p<0.05$). There were no significant main effects of object location or sentence direction. Again, there were too few items to expect an item analysis using ANOVA to yield significant results, but looking at them individually (Table. 6), we see that almost all of the sentences with down-associated subject nouns yielded faster categorization when the subsequent object appeared in the upper part of the visual field. Interestingly, the one exceptional sentence in this group, *The submarine fired*, might be construed as encoding upwards movement. That is, when submarines fire ballistic missiles rather than torpedoes, they typically fire upwards. The sentences with up-related subject nouns showed the opposite tendency, as predicted. Namely, the majority yielded faster response times to the categorization task when the object appeared in the lower part of the screen.

3.3. Discussion

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

Note that unlike the sentences with verbs denoting upwards or downwards 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 (since the sentences in Experiment 2 did not encode upwards or downwards motion so much as up or down location). Either of these accounts would predict the asymmetry to disappear in this second study. And 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, while 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 towards or away from the participant (Zwaan et al.,

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

4. Experiment 3: Metaphorical sentences

Language about motion in a direction, or about objects located in a given location, yielded significant interference on a visual perception task in the first two studies. To investigate whether this effect was the result of lexical or sentential interpretation, we performed a third experiment

23

testing whether sentences with motion verbs but which denoted no literal motion would also interfere with object categorization.

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

(7)   a.   *The drainpipe <u>climbs</u> up the back wall of the house.*

       b.   *Starting at the house, the fence <u>drops</u> down quickly to the ocean.*

       c.   *Investors sold stocks lower Tuesday as oil prices <u>climbed</u> above $51 per barrel.*

       d.   *Mortgage rates <u>dropped</u> further below 6 percent this week.*

The interpretation processes involved in understanding figurative language have been a matter of significant research and debate. Some work has demonstrated that language users access internal representations of space and motion when performing reasoning tasks about abstract concepts understood metaphorically in terms of these concrete notions (Gibbs et al., 1997; Boroditsky, 2000; Boroditsky & Ramscar, 2002). Moreover, there is limited evidence that processing connected discourse using metaphor proceeds most quickly when conventional metaphorical expressions are used (Langston, 2002). However, we do not yet know whether simply processing metaphorical motion language makes use of spatial representations. Critically, if the effect observed above in the first two experiments is simply lexical, or if figurative language yields the same visual imagery that literal language does, then we should expect to see no difference when the same experiment described above is conducted with figurative upwards or downwards motion sentences, rather than literal ones. However, if the effect is due to the interpretation of the sentence - where a participant mentally simulates the described scene - and does not simply result from the lexical semantics of constituent words (and if figurative language differs in some ways

24

from literal language interpretation) then we expect to see a significant decrease 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 with figurative ones.

4.1 Method

All the motion verbs used in the first study on literal sentences above (section 2) can also be used to describe changes in quantity or value of entities that do not have physical height, such as oil prices or mortgage rates (7c,d). Thus, to create metaphorical sentences, we used subjects such as *rates* and *prices* along with the same motion verbs used in the first experiment to produce metaphorical sentences. The sentences were normed as described in Section 2.2. The up and down metaphorical sentences showed no significant difference in reaction time $F(1,27)=0.07$, $p=0.793$ or in meaningfulness rating: $F(1,27) = 0.969$; $p=0.334$ (Fig. 2). The nouns in metaphorical up versus down sentences were not rated different in upness: $F(1,28)=1.209$; $p=0.281$, or in downness: $F(1,27)=0.003$; $p=0.954$, while the verbs were, as seen in Table 1.

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

4.2. Results

As can be seen from Table 3, by contrast with the literal verb and noun sentences, there was no significant interaction effect between sentence direction and object location with the metaphorical sentences ($F(1,58)=0.425$; $p=0.517$) Nor were there significant main effects of object location or sentence direction. The analysis of items (Table 7) reveals the same pattern: More sentences in the down condition yielded faster response times when the object was in the lower half of the visual

field, and the reverse was true for metaphorical up sentences. Both of these tendencies were the reverse of the predicted direction of the Perky effect.

4.3. Discussion

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

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

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

5. Experiment 4: Abstract sentences

This final experiment tested whether abstract sentences produce locationally selective interference on a visual categorization task. Our abstract sentences, like the metaphorical sentences in Experiment 3, denoted changes in quantity, but did so using verbs that did not also have a concrete meaning denoting change in height (verbs such as *increase* and *wane*). Embodied accounts of conceptual representation and language understanding (Lakoff, 1987; Barsalou, 1999;Glenberg & Robertson, 2000) argue that all concepts, whether concrete or abstract, are ultimately grounded in terms of embodied individual human experience in the world. The grounding of concrete concepts can be straightforwardly accounted for in terms of the perceptual, motor, and perhaps even affective content of experiences an agent has when dealing 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 (Lakoff & Johnson, 1980;Grady, 1997). Indeed, much of the time 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.   The verbs express a change in quantity, either an increase, such as *increase,* and *double*, or a decrease, such as *decrease* and *lessen*. Abstract verbs 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 (8).

(8)   a.   *The figures doubled.*

        b.   *The percentage decreased.*

Since the abstract verbs used here do not denote any literal upwards or downwards 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 ($F(1,28)=86.489$; $p<0.001$) and downness rating ($F(1,27)=149.777$; $p<0.001$). By contrast, the nouns in abstract up versus down sentences were not rated differently in upness: $F(1,28)=0.026$; $p=0.874$) or in down-ness: $F(1,27)=0.074$; $p=0.787$ (Fig. 1). Abstract sentences in the two conditions showed no significant difference in the reaction times: $F(1,28)=1.539$; $p=0.225$; nor in the meaningfulness ratings: $F(1,28)=0.007$; p=0.935.

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

5.2. Results

By contrast with the literal up and down sentences, the means for the abstract sentences show no interference effect (Table 3). Indeed, a participant analysis of reaction times following abstract sentences showed no significant interaction of sentence direction with object location ($F(1,63)=0.126$; $p=.724$). There were no significant main effects of sentence direction or object location, either. The individual items in the abstract condition did not display the polarization seen in the responses to individual items in the literal sentences in Experiments 1 and 2. As many abstract down sentences (three of five) yield longer response times whether the object is displayed in the upper or the lower part of the visual field.

5.3. Discussion

Despite being systematically associated with upness or downness, the abstract verbs used in this experiment did not yield selective interference on the object categorization task. This provides further evidence that the outcomes of the Experiment 1 and Experiment 2 did not result simply from lexical attributes of the constituent words in the sentences – something like a [+UP] or [+DOWN] feature. The abstract up verbs were strongly up-associated, and the abstract down verbs were strongly down-associated, at least as measured by the norming data, and yet these aspects of their semantics were not sufficient for them to interfere with visual object categorization. There is a straightforward explanation for the presence of an interference effect in the first two studies and its absence in the last two. Namely, the scenes described by the first two involved actual action in one direction or the other, while those described by the last two did not. It would thus seem to be the construction of a mental representation of the described scene, rather than purely lexical semantics, that drives the measured interference effect.

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

Despite this variability in experimental results, it has been widely suggested that we base abstract thought and language on concrete thought and language (Lakoff, 1987; Barsalou, 1999; Barsalou & Wiemer-Hastings, 2005). This study shows that it is not straightforwardly the case that a particular abstract domain is processed exactly as the concrete domain it is supposedly related to. Of course, this should not be particularly surprising. If individuals actually enacted mental imagery in order to understand abstract language where that imagery was not qualitatively different from imagery that resulted from literal language processing, this would be a confusing state of affairs during language comprehension indeed. And since we know that in understanding language, people are not prone to confusing changes in quantity of abstract numbers with change in height of physical objects, the processing of these different domains must differ in some ways.

It remains to be seen exactly what processes underlie abstract language understanding, but the absence of an interference effect observed here does not imply that the embodied account for abstract language understanding and abstract concept grounding is incorrect. There may be other factors that obscure a measurable interference effect with abstract sentences, entertained in Section 6, below. A clear result of this final experiment, however, is that where Richardson et al.'s (2003) earlier work showed that abstract sentences yield interference effects on categorizing objects in the same axis, we found no effect of abstract sentences on categorizing objects in the same location. Below, we consider several possible explanations for this difference, such as differences in how abstract the abstract stimuli were, and differences in the detail of mental imagery driven by concrete versus abstract language.

6. General Discussion

Processing sentences denoting events that would tend to take place in a particular part of a perceiver's visual field yields interference on actually using the same part of the real visual field, as measured by decreased performance in an object categorization task. This is true whether the location of the event is indicated by a verb (Experiment 1) or a sentential subject (Experiment 2). However, having an up- or down-associated lexical item in a sentence does not suffice to produce interference. The sentence must literally encode a scene involving the relevant location in the visual field, as metaphorical uses of motion verbs (Experiment 3) and abstract verbs that are nonetheless associated with upness or downness (Experiment 4) yield no significant interference effect. We can conclude from this that it is not lexical priming that yields the interference but rather the performance of mental imagery corresponding to the meaning of an utterance.

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

Second, the results from the four studies demonstrate a clearly longer latency for reactions to objects following the abstract sentences than to any of the other sentence types. A possible explanation for this that the abstract sentences were simply harder to process, thus resulting in slower reactions to the following stimulus. However, as we see in Fig. 1, metaphorical sentences, and not abstract sentences, were read most slowly, so unless the reading times in the norming study

differed significantly from the processing time in the main experiments, this does not explain the slow reactions to object categorization following abstract sentences.

A final remaining pair of issues is why abstract sentences showed significant interference in Richardson et al. (2003) but not in the current study, while literal sentences did show significant effects in the current study, while they did not appear to do so – albeit in statistically unlicensed post hoc tests – in Richardson et al. (2003).[3] Though the two studies differed along several dimensions, the most relevant one is the assignment of sentences to different conditions. The original study took upwards- and downwards-directed sentences as belonging to the same condition, and categorized all responses to objects appearing either in the upper or the lower part of the screen as belonging to the same condition - thus being only specific to the axis of concrete or abstract motion. By contrast, the current study pulled apart the up and down conditions in sentences and object responses. This offers a straightforward explanation for the difference in responses to literal sentences in the two experiments.

Given that we have seen in the current work that literal up sentences interfere with visual processing in the upper part of the screen, and down sentences interfere with the lower part of the visual field, it is not at all surprising that grouping all these responses together (as was done in Richardson et al., 2003) would eliminate any effects. After all, up sentences (possibly about half of the sentence in the vertical condition) would result in slower responses to objects in the upper part of the screen (half of the objects in that condition), while the down sentences (the remaining sentences in that same condition), would interfere with the other half of the object stimuli - those in the lower position. The two effects would cancel each other out, resulting in no significant effect. By contrast, the current study, which investigated not just axes but more particularly locations along those axes did not see such effects obscured, and the results were thus clearly significant.

There are several candidate explanations for abstract sentences showing a significant interference effect in the original study (Richardson et al., 2003), but not the current study. The first is based on this same structural difference between the experiments, focusing on locations or collapsing them into axes. We have seen in the current work (Experiment 4) that abstract sentences do not yield interference effects, which differs from the results reported in Richardson et al. (2003). The main difference was again that the earlier study took the axis (vertical or horizontal) as the main dimension of difference, while our investigation was specific to location. Something like the opposite of the explanation provided for the literal sentence difference may apply to the abstract sentences. Namely, perhaps, as Richardson et al. have shown, abstract sentences do trigger mental imagery, but imagery that is not specific to particular locations so much as to axes. That is, abstract language imagery is less spatially precise, while still retaining an imagistic component. If this is the case, then we should predict that when up and down are collapsed together and the entire vertical axis is treated as a condition, then abstract language should yield measurable interference effects. However, a study like Experiment 4, in which objects located in the upper and lower regions are placed in separate conditions, should show no such interference, because the abstract motion sentences are not incompatible with any of the presented objects, all of which appear in the vertical axis.

There is some support for this account in evidence that axes and specific locations are represented distinctly in the human cognitive system, as suggested by Logan and Sadler (1996). Carlson-Radvansky and Jiang (1998) have shown that individual words like *above* may activate an entire axis, presumably as contrasted with location-specific words like *up*. McCloskey and Rapp (2000) have similarly shown that axis and direction can dissociate in particular neurological disorders. A participant they studied had lost the ability to ballistically reach for targets (thus, lost specific location), but preserved the correct axis along which the object was located. Similarly, Landau and Hoffman (2005) have shown that children with Williams Syndrome have difficulty

with direction but not axis of orientation. Thus, there is good evidence from relevant studies that object location may be represented separately from axis of orientation, and as such the two different systems might be available to be recruited by concrete versus abstract language processing.

This explanation, global interference in the abstract condition, would also explain the finding, mentioned above, that abstract sentences yielded considerably longer object categorization times overall than any of the other sentence types. If abstract language about upwards or downwards motion activates imagery along the entire vertical axis, than any abstract up/down sentence should yield an increased latency for object perception anywhere along the vertical axis - meaning slower reaction times in all four combinations of sentence direction and object location.

A second explanation for the absence of an effect with abstract sentences in our Experiment 4, but the presence of such an effect in Richardson et al.'s work, relies on differences in the abstractness of the stimuli in the two studies. In the original work, abstract sentences included verbs rated as abstract in the MRC Psycholinguistic database. This selection method may have inadvertently resulted in a small number of relatively concrete verbs; perusing the verbs in their study yields several candidates, like *argue*, *rush*, *give*, and *rest*. These verbs were combined with arguments, which were mostly very concrete - sentential subjects always denoted people, like *the storeowner, the husband*, or *the jogger*. The combination of even relatively abstract verbs – like *want* – with concrete arguments – like *the child* and *the cake* – results in sentences that could easily yield mental imagery of concrete situations. In this example, an imagined scenario in which a child wants cake might involve a child looking covetously at some cake in a spatial arrangement that is probably horizontal. To summarize this line of reasoning, abstract sentences in the original study contained some linguistic elements that might have made the scenes they described concretely imageable, and those images might have been responsible for the interference effect observed with these abstract sentences.

By contrast, abstract sentences in the current study (Experiment 4) were more abstract. All verbs (in Table 8) denoted change in quantity (though some, such as *expand* are inevitably somewhat concrete as in Richardson et al.'s study). But the nouns in the sentences are all abstract, describing quantitative measures like *quantity*, *ratio*, and *measures*. As a result, it is subjectively more difficult to imagine a concrete scene in which the scenes these sentences describe would be grounded than it is for the abstract sentences in the original study. This could be responsible for the difference findings in the two studies – perhaps abstract language only yields measurable imagery effects when it is straightforwardly interpreted as referring to spatially concrete scenes.

We tested these two explanations through a follow-up study with 50 participants that pitted the two explanations against each other. What we did was to create new abstract up and down sentences using the same template as in Experiment 4, so that we had twice as many up- and down-oriented abstract sentences of the same high degree of abstractness. We used the exact same methodology as in Experiment 4, except that these critical sentences were now followed by objects appearing not only in the upper and lower parts of the screen, but also on the right and left. If we found an effect of axis but not quadrant – that is, if abstract sentences yielded slower response times to object categorization in the upper and lower parts of the screen than in the left and right parts – this would replicate Richardson et al.'s (2003) findings and support the hypothesis that abstract sentences are simulated with less detail than concrete ones. However, if there was no simulation effect at all for these abstract sentences by axis (and if, as in Experiment 4, there was none by quadrant), then this would support the view that our sentences were more abstract than Richardson et al's, and that this difference in stimuli was what yielded the absence of an effect in our study.

As it turned out, the latter was what we found: abstract up and down sentences yielded responses to objects in the vertical and horizontal axes that were statistically indistinguishable (repeated- measures ANOVA by participants: $F(1,49)=1.68$, $p=0.20$), and as in Experiment 4,

there was no interaction effect between sentence direction and up or down object location (repeated-measures ANOVA by participants: $F(1,49)=0.41$, $p=0.53$). These results suggest that it was differences in the concreteness of the stimuli used in the original paper and the present work that produced different imagery effects. The absence of an effect in both our Experiment 4 and the follow-up reported here could be due to greater variation in the specific imagistic realization of sentence processing when the language has few concrete anchors. For example, abstract thought about quantity might sometimes use vertical visual imagery, but other times other concrete domains like physical volume, size, or mass. It appears that abstract language is capable of driving mental imagery, as long as it describes scenes that are consistently imageable in consistent concrete terms.

It is worth noting that this explanation still does not fully explain the difference between the two studies. There still must be differences (whether these be theoretically interesting or simple artifacts of stimuli selection) between the concrete and abstract sentences in the original study by Richarson et al. (2003). Otherwise, the abstract sentences they used – which we have argued to be somewhat concrete due to their sentential arguments – should have yielded no effect, like the concrete sentences. This leaves open the possibility that both explanations outlined above are true – that not only do abstract sentences yield stronger imagery effects when they include tangible subjects and objects, but in a addition the imagery they produce is less spatially precise than that performed for concrete sentences.

The findings reported in the foregoing studies provide new evidence suggesting that understanding motion language leads individuals to activate internal simulations of the described scenes. While the selective interference of language processing on visual perception does not imply that simulation is required for language understanding, it does imply that it is unconscious and automatic. Various authors have suggested different roles for the construction of a mental simulation on the basis of language, using detailed modal knowledge. One critical role of imagery is

to produce detailed inferences (Narayanan, 1997), which can both allow an individual to gain a rich notion of the utterance's content, such as a situation model of the described scene (Zwaan, 1999), as well as to prepare the individual to understand future utterances, or to respond relevantly. Second, the construction of a mental simulation prepares the individual for situated action (Bailey, 1997; Barsalou, 1999b; Glenberg & Kaschak, 2002). Finally, some language may be disambiguated only through the performance of imagery (Bergen & Chang, 2005).

Indeed, various theories of language rely heavily on perceptually and motorically grounded representations as the backbone for the language understanding process. Of particular note, Kashak and Glenberg (2002) argue that language understanding proceeds through the meshing of simulation constraints from language, and the subsequent mental simulation of afforded actions, to prepare for situated responses. Zwaan (1999, 2004) has argued similarly that language comprehension proceeds through the construction of modal mental models, and Barsalou (1999) has suggested that language hooks into simulators - systematic patterns of reactivation of representations of perceptual and motor experiences. What all these models share is a recognition of the importance of mental simulation in the process of language understanding. However, none of them are actual theories of how the individual linguistic items that make up an utterance directly produce a mental simulation, especially given the complexities of linguistic structure (although Kashak and Glenberg (2002) have made some progress with regard to how 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 the precise ways in which language triggers simulation, and what aspects of simulation it triggers. Kashak 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 (Langacker, 1987; Lakoff, 1987; Talmy, 2000). Though cognitive linguistic work is based largely on introspection, 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.

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). 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 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.

Visual interference effects produced by linguistic input are reliable and replicable, in a number of methodological permutations. These findings as a whole can be taken as 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 (Glenberg & Kaschak 2002, Bergen & Chang 2005). 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

References

Barnett, V., & T. Lewis. (1978.) *Outliers in Statistical Data.* New York: John Wiley & Sons.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-609.

Barsalou, L.W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher and R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129-163). New York: Cambridge University Press.

Bergen, B. & N. Chang. (2005). Embodied construction grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive grounding and theoretical extensions* (pp. 147-190). John Benjamins.

Bergen, B. (in press). Experimental methods for simulation semantics. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, and M.J. Spivey (eds.) *Methods in Cognitive Linguistics.* Ithaca. John Benjamins.

Bergen, B., N. Chang, & S. Narayan (2004). Simulated Action in an Embodied Construction Grammar. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society.*

Bergen, B., S. Narayan, & J. Feldman (2003). Embodied verbal semantics: evidence from an image-verb matching task. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society.*

Boroditsky, L. (2000). Metaphoric Structuring: Understanding time through spatial metaphors. *Cognition*, *75(1)*, 1-28.

Boroditsky, L. & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, *13(2),* 185-188.

Bower, G. & Morrow, D. (1990). Mental models in narrative comprehension. *Science*, 247, 44-48

Carlson-Radvansky, L.A., & Jiang, Y. (1998). Inhibition accompanies reference frame selection. *Psychological Science*, *9*, 386-391.

Ehrsson, H.H., S. Geyer, & E. Naito (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part specific motor representations. *Journal of Neurophysiology, 90,* 3304–3316.

Gibbs, R., Bogdonovich, J., Sykes, J., & Barr, D. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language,* 37, 141-154.

Glenberg, A. & D. Robertson (2000). Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language, 43,* 379-401.

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

Johansson, R, Holsanova, J. and 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.

Kaschak, M. & A. Glenberg (2000). Constructing Meaning: The Role of Affordances and Grammatical. Constructions in Sentence Comprehension. *Journal of Memory and Language 43,* 508–529.

Kaschak, M.P., Madden, C.J., Therriault, D.J., Yaxley, R.H., Aveyard, M., Blanchard, A.A., & Zwaan, R.A. (2005). Perception of motion affects language processing. *Cognition* 94(3): B79-B89.

Kersten, A. (1998). A division of labor between nouns and verbs in the representation of motion *Journal of Experimental Psychology: General,* 127(1), 34-54.

Kosslyn, S. (1980). *Image and Mind.* Harvard University Press.

Kosslyn, S., G. Ganis, & W. Thompson (2001). Neural foundations of imagery. Nature Reviews *Neuroscience, 2,* 635-642.

Lakoff, G. (1987). *Women, fire, and dangerous things.* Chicago: Chicago University Press.

Lakoff, G. & M. Johnson (1980). *Metaphors we live by.* Chicago: Chicago University Press.

Landau, B., & Hoffman, J.E. (2005). Parallels between spatial cognition and spatial language: Evidence from Williams syndrome. *Journal of Memory and Language.* 53(2), 163-185

Langston, W. (2002). Violating orientational metaphors slows reading. *Discourse Processes*, 34(3), 281-310.

Lindsay, S. (2003). *Visual priming of language comprehension.* Unpublished University of Sussex Master's Thesis.

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: M.I.T. Press.

Matlock, T. (2004a). The conceptual motivation of fictive motion. In G. Radden and R. Dirven (Eds.), *Motivation in grammar.* (pp. 221-248). Amsterdam: John H. Benjamins.

Matlock, T. (2004b). Fictive motion as cognitive simulation. *Memory & Cognition* 32, 1389-1400.

Matlock, T. & Richardson, D.C, (2004). Do eye movements go with fictive motion? *Proceedings of the Twenty-sixth Annual Meeting of the Cognitive Science Society.*

McCloskey, M., & Rapp, B. (2000). Attention-referenced visual representations: Evidence from impaired visual localization. *Journal of Experimental Psychology*, 26(3), 917-933.

Narayan, S., B. Bergen, & Z. Weinberg (2004). Embodied Verbal Semantics: Evidence from a Lexical Matching Task. *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society.*

Nyberg, L., K.-M. Petersson, L.-G. Nilsson, J. Sandblom, C. Åberg, & M. Ingvar. (2001). Reactivation of motor brain areas during explicit memory for actions. *NeuroImage, 14*, 521-528.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510-532.

Richardson, D.C., & Matlock, T. (in press). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition.*

Richardson, D.C., Spivey, M.J., & Cheung, J. (2001) Motor representations in memory and mental models: Embodiment in cognition. *Proceedings of the Twenty-third Annual Meeting of the Cognitive Science Society*: 867-872, Erlbaum: Mawhah, NJ.

Richardson, D. C., Spivey, M. J., McRae, K., & Barsalou, L. W. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, 27, 767-780.

Spivey, M. & Geng, J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, 65, 235-241.

Stanfield, R.A. & Zwaan, R.A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12, 153-156.

Zwaan, R.A., Madden, C.J., Yaxley, R.H. & Aveyard, M.E. (2004). Moving words: dynamic representations in language comprehension. *Cognitive Science*, 28, 611-619.

Zwaan, R.A., Stanfield, R.A., Yaxley, R.H. (2002). Do language comprehenders routinely represent the shapes of objects? *Psychological Science*, 13, 168-171.

Table 1. Results of norming studies in which participants rated nouns and verbs on upness and downness. Scale was 1 through 7.

	Nouns		Verbs	
n = 28	Up Avg (s.d.)	Down Avg (s.d.)	Up Avg (s.d.)	Down Avg (s.d.)
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)

Table 2. Results of norming studies in which participants read sentences and rated them on maningfulness. Meaningfulness scale was 1 to 7.

n = 28	Mean RT (s.d.)	Mean Meaningfulness (s.d.)
Experiment 1		
Down (verb)	1515 (631)	6.16 (0.81)
Up (verb)	1844 (813)	5.81 (0.96)
Experiment 2		
Down (noun)	1691 (828)	6.31 (0.88)
Up (noun)	1554 (624)	6.48 (0.88)
Experiment 3		
Down (metaphor)	1970 (832)	5.41 (1.04)
Up (metaphor)	2011 (1036)	5.59 (0.92)
Experiment 4		
Down (abstract)	1932 (875)	6.13 (0.80)
Up (abstract)	1811 (806)	6.12 (0.75)

Table 3 - Mean RT for object categorization in upper and lower quadrants of the screen.

	Object in lower quadrant - mean RT (s.d.) (s.e.)	Object in upper quadrant - mean RT (s.d.) (s.e.)	Difference (RT)
Experiment 1			
Down (verb)	551 (255) (32)	542 (240) (30)	+9
Up (verb)	526 (205) (26)	603 (270) (34)	-77
Experiment 2			
Down (noun)	550 (221) (28)	506 (245) (20)	+46
Up (noun)	508 (218) (30)	526 (247) (22)	-18
Experiment 3			
Down (metaphor)	516 (283) (23)	532 (228) (25)	-16
Up (metaphor)	535 (235) (24)	531 (240) (24)	+4
Experiment 4			
Down (abstract)	589 (230) (29)	575 (222) (28)	+14
Up (abstract)	593 (268) (33)	600 (317) (40)	-7

Table 4 - Mean RT for object categorization in upper and lower quadrants of the screen for Up and Down Sentences in Experiment 1, by sentence

Up Sentences	Object Up	Object Down
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 Sentences	Object Up	Object 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

Table 5 - Mean RT for object categorization in upper and lower quadrants of the screen, for the first half of Experiment 1 only

	Object in lower quadrant - mean RT (s.e.)	Object in upper quadrant - mean RT (s.e.)	Difference (RT)
First Half Only			
Down (verb)	604 (34)	561 (28)	+43
Up (verb)	593 (29)	626 (39)	-33

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

Noun Down Sentences	Object Up	Object 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 Sentences	Object Up	Object Down
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**

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

| Metaphorical Down Sentences	Object Up	Object 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 Sentences	Object Up	Object Down
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**

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

Abstract Down Sentences	Object Up	Object 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 Sentences	Object Up	Object Down
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**

Figure 1 - Mean reading time for each of the sentence types
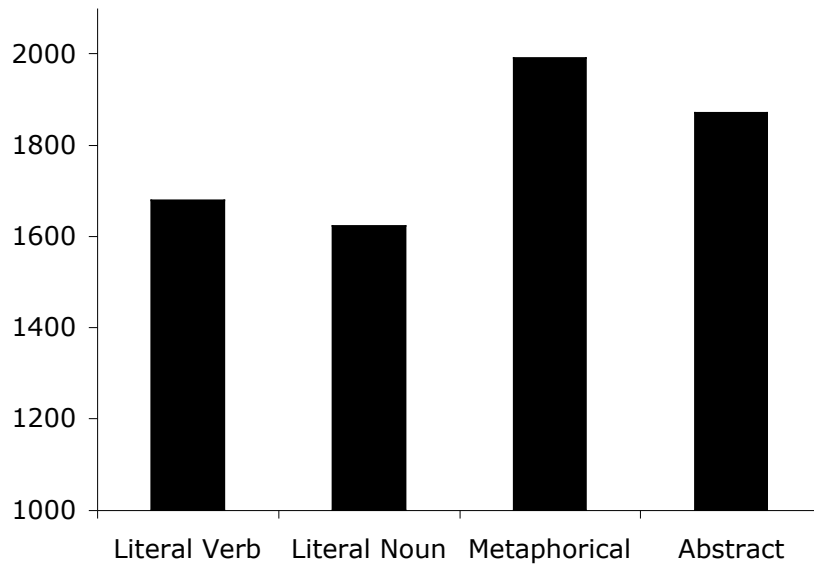
[1] The relatively small number of sentences of each type could in principle be remedied by using the words *up* and *down* in sentences. We chose to avoid these words for several reasons. First was the possibility that participants would recognize these recurring words in the experiment and guess its purpose. We were also concerned with potential direct effects of the words *up* and *down* on participants' responses. For example, seeing those words might result in participants orienting overt attention to that part of the visual field, which would counteract the expected effect. Conversely, the presence of those words might be argued to be responsible themselves for any observed effects, rather than the interpretation of the sentence as a whole (which we tested contrasting experiments 1 and 3.).

[2] Replacing outliers with values at a set distance from the subject's mean is also known as "Windsorizing" (Barnett & Lewis, 1978) and is commonly used in sentence processing research. While it may increase power in a small set of restricted cases, it globally does not affect results of statistical analyses (Ratcliff, 1993). We chose to Windsorize, rather than eliminate outliers, due to the small number of items in each condition.

[3] In a footnote, the authors report a sizeable effect (F=6.92) for abstract sentences but none for concrete ones (F=0.59). This difference could in principle be due to reduced statistical power in considering only half of the items, but at the very least it leads us to conclude that the effect was not as strong for concrete as for abstract sentences.