2006

# Investigating the Simulation of Event Locus during Online Comprehension

Richard H. Yaxley

THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

# INVESTIGATING THE SIMULATION OF EVENT LOCUS DURING ONLINE COMPREHENSION

By

RICHARD H. YAXLEY

A Dissertation submitted to the
Department of Psychology
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Fall Semester, 2006

The members of the Committee approve the Dissertation of Richard H. Yaxley defended on September 15, 2006.

Rolf A. Zwaan
Professor Directing Dissertation

Leonard L. LaPointe
Outside Committee Member

Michael P. Kaschak
Committee Member

Colleen M. Kelley
Committee Member

Jon K. Maner
Committee Member

The Office of Graduate Studies has verified and approved the above named committee members.

# TABLE OF CONTENTS

# LIST OF FIGURES

# ABSTRACT

Recent evidence in support of embodied cognition suggests that cognition in general, and language comprehension in particular, calls upon sensorimotor traces to mentally simulate described situations. Although the extent to which such experiential traces are integral to linguistic representations remains an open theoretical question, recent eye-tracking studies have demonstrated an immediate interfacing of linguistic and visual representations during understanding. Three eye-tracking experiments were performed that investigated the extent to which the processing of described events involves the anticipatory allocation of visual attention. All three experiments demonstrated that visual attention is allocated to regions within entities equivalently despite the implicit focus of the described event. The results did not provide supporting evidence for the hypothesis that language processing activates context-specific regions during mental simulation. These findings are discussed in the light of evidence that shows that comprehenders' ability to disambiguate meaning and establish reference with external entities can be influenced by linguistic and environmental context.

# CHAPTER 1

# INTRODUCTION

Language is a tool used to convey meaning. The intended meaning may refer to situations in the immediate environment or situations displaced in space or time. Therefore, understanding the meaning behind language involves deciphering the relevant correspondence between linguistic expressions and the objects and events to which those expressions refer. Although people perform this skill seemingly effortlessly, relatively little is known about the cognitive mechanisms that support this behavior. Recent empirical investigations into cognitive processing suggests that the mental representations that underly language comprehension are actually analogues of sensorimotor activation. In other words, the lower-level representations that are involved in perception and action are also involved in higher-level conceptual processing (Barsalou, 1999). Thus, language-based mental representations are not conceptual abstractions derived from strings of words. Rather understanding the meaning behind language is inherently grounded in one's embodied experiences with objects and events (for a review see Wilson, 2002; Harnad, 1990).

Barsalou (1999) suggests that the cognitive processes that support understanding recruits these sensorimotor representations in a process of mental simulation. Simulations are thought to involve the partial reactivation of neural patterns that occur during stimulation. Several studies have demonstrated that language processing invokes perceptual activation that resembles the patterned activation involved in direct experience (Glenberg & Kaschak, 2002; Kaschak et al., 2005; Zwaan et al., 2002; Stanfield & Zwaan, 2001; Zwaan & Yaxley, 2003; Yaxley & Zwaan, in press). This select activation is indicative of a system that uses linguistic cues to reactivate the experiential traces associated with described referents (Barsalou, 1999; Glenberg, 1997; Goldstone, 1998; Zwaan, 2004). Furthermore, Zwaan & Kaschak (in press) posit that such inferences emerge from of an ongoing process of prediction and resonance that ultimately enables comprehenders to make knowledge-based inferences about decribed situations and anticipate events. However, much remains unknown about by the intricacies of mental simulation and the degree to which comprehenders develop such expectations. For instance, it is unknown if language-based mental simulations invoke only the activation of object-level representations, or if the simulations also invoke more fine-grained activation of knowledge about objects. The present study was designed to investigate this question by testing if the comprehension of described events involves the activation of context-specific knowledge associated with the different functional regions within objects.

In order to understand how closely mental simulations map onto direct experience, it is also important to determine how simulations evolve over time. Most of the studies that have found supporting evidence for simulations have relied on participants' conscious judgments about stimuli (e.g., picture verification). It is therefore unclear if comprehenders' simulations evolve automatically while linguistic content is being processed, or if the activation develops as part of more conscious strategizing by the participants. One way to investigate this issue is by looking at the behavior associated with online language processing. For example, several eye-tracking studies have demonstrated that linguistic processing has an immediate and routine influence on one's ability to establish reference with visual entities (Cooper, 1974; Tanenhaus et al., 1995). Using what has become known as the 'visual-world' paradigm, several studies have demonstrated that perceptual and conceptual overlap between linguistic content and environmental context integrate in real-time to constrain understanding. Thus, participants utilize the environment in which they are situated to support ongoing comprehension (O'Regan, 1992; Henderson & Ferreira, 2004). This online integration of linguistic and visual representations also demonstrates that simulations are not static constructions, but evolve incrementally as linguistic input is processed (e.g., Altmann & Kamide, 2004).

This thesis will briefly review evidence for the mental simulation of linguistic expressions in Section 2. Section 2.1 will discuss how linguistic simulations evolve as linguistic content is processed. Three novel eyetracking experiments are described in Section 3, and each is evaluated to determine if language comprehension activates knowledge based on the events verbs denote and how this influences visual attention. Finally, a summary of this thesis and a discussion of future work is provided in Section 4.

# CHAPTER 2

# LANGUAGE-BASED SIMULATIONS

Language comprehension has been shown to call upon experiential knowledge that influences one's understanding and interaction with the world. For example, recent studies have demonstrated that comprehension evokes context-specific knowledge (e.g., Zwaan et al., 2002). The context-specific knowledge that is activated is hypothesized to stem from the mental simulation of referential events (Barsalou, 1999). The simulation of described events recruits the sensorimotor representations that are associated with one's experiential traces of specific situations. Because events activate a finite set of sensorimotor representations, the simulation of those events should activate the same set, or at least a subset, of experiential traces (Zwaan, 2004). Simulations are therefore capable of selectively activating sensorimotor representations that are linked with contextually-appropriate properties of situations (Barsalou, 2003). What follows is a brief review of how this context-specific activation of knowledge influences several aspects of semantic processing.

A classic memory study demonstrates clearly how language can impact one's retrieval of observed events (Loftus & Palmer, 1974). In this study, participants viewed a film on traffic safety that included an automobile accident. Subsequently, they were asked to estimate the velocity of the cars when they [contacted/hit/smashed] each other. The observers' speed estimates were shown to be directly influenced by the intensity of the verb used in the question, such that 'contacted' elicited the lowest speed estimates and 'smashed' elicited the highest. When the participants were questioned later about the observed accident, their memory for the collision was influenced by the verb used in the initial question session, such that the more intense verb 'smashed' elicited more memories details commonly associated with auto collisions, but that were actually not present in the film (e.g., broken glass on the road). These findings suggest that memory for directly observed events is influenced by language-based mental simulations. Furthermore, it appears that language can invoke mental simulations that generate inferences, which then influence memory by integrating with the sensorimotor representations associated with actual experience. This study provides a striking example of how linguistic cues can, unconsciously, be integrated into one's mental representations. But how?

Recent evidence suggests that language comprehension involves the reactivation of experiential traces by invoking mental simulations of the described referents. These simulations involve the patterned sensorimotor activation associated with the experiences of the referential objects and events described. Studies have explored the extent to which language-based simulations are analogues of direct experience by investigating the integration of linguistic processing with perceptual processing. The simulations reactivate contextually-specified properties associated with the aggregate of direct experience. This sensorimotor activation appears to influence comprehension in several ways. For example, such activation seems to enable comprehenders to make spontaneous inferences about perceptual and motoric properties of referential events. In many of these studies, focus has been placed on the extent to which linguistic context invokes visuo-spatial properties and subsequently influences the interpretation of objects.

For example, language processing has been shown to evoke the activation of fundamental object properties like shape and orientations. Zwaan and colleagues have demonstrated in several studies that comprehension shows an immediate effect on picture verification. For example, Zwaan et al. (2002) had participants read a sentence such as 'The park ranger saw the eagle in the [sky/nest]'. Participants then had to judge whether a pictured eagle was mentioned in the previous sentence. Critically, the target object (eagle) was depicted in a shape (wings outstretched or drawn in) that was congruous with the implied situation or not. Picture verification times were influenced by the object's described location such that judgments were longer when the shape of the depicted eagle did not match the shape implied in the sentence (i.e., wings outstretched for 'in the sky', or wings withdrawn for 'in the nest'). A similar study also demonstrated that comprehension activates orientation information. Participants read sentences such as 'The carpenter hammered the nail into the [floor/wall]'. Critically, the pictured nail that was judged in either a vertical or horizontal orientation. As in the Zwaan et al. (2002), picture verification times were longer when the orientation of the pictured nail was incongruent with the orientation implied in the description (e.g., a vertical nail after 'into the wall') than for those that were congruent. Together, these findings suggest that language comprehension involves the routine activation of contextually-specified shape and orientation information.

The manner in which language is presented also influences semantic processing. Zwaan & Yaxley (2003) demonstrated this by having participants judge the semantic association of word-pairs that referred to two different parts of an object (e.g., 'branch' and 'root'). The word-pairs named parts of objects that have a canonical spatial orientation (e.g., trees typically have branches above their roots). The word-pairs were presented to the participants with one word above the other so that their relative placement either matched or mismatched their canonical layout (see 2.1).

| Congruous | Incongruous |
|---|---|
| BRANCH<br>+<br>ROOT | ROOT<br>+<br>BRANCH |

Figure 2.1: Relative orientation of words affects semantic relatedness judgments.

Participants were slower to judge the word relations when the spatial layout did not match their canonical spatial orientation (i.e., when 'root' was presented above 'branch'). This finding was interpreted as further evidence for the simulation hypothesis. The semantic comparison of the names that refer to parts of an object involves the activation of experiential traces associated with the referential entity. The semantic judgments apparently involved the activation of the spatial representation of the referents layout. Although there was no instruction or linguistic information to influence how the word-pairs should be construed, spatial context proved influential in semantic processing.

It has also been shown that environmental context influences how one processes instructions. Carlson & Kenny (2006) and colleagues demonstrated that the functional relationships between objects influence how one interprets spatial terms. They presented participants with pairs of pictured objects and instructed them to place one object next to the other (e.g., 'Place the bottle opener next to the beer bottle'). When participants placed an object that was functionally related to the target (e.g., beer bottle and bottle opener), they placed the located object (bottle opener) next to the functionally relevant region of the reference object (the bottle's cap). However, when there was no functional relationship between the located and reference objects, the located object was placed near the middle of the bottle. Thus, the spatial relationships between unrelated objects appears to be dependent on geometrical properties, whereas the spatial relationships between functionally related objects are influenced by their functional associations.

The demonstrations discussed above provide strong evidence that mental simulations are involved in comprehension. However, it remains unclear to what extent such simulations recruit sensorimotor representations and how these representations are linked together to support comprehension. It seems that linguistic context enables the comprehender to disambiguate meaning by selectively activating the properties of referential objects. This selective activation then results in the spontaneous development of inferences. Zwaan & Kaschak (in press) posit that such inferences emerge from of an ongoing process of sensorimotor resonance that ultimately enables comprehenders to anticipate events and prepare for action. The extent to which these language-based simulations enable such anticipatory responses is thought to be dependent on the degree of correspondence between linguistic cues and experiential representations. Furthermore, this correspondence is not static, but rather evolves over time as linguistic cues incrementally activate the experiential traces associated with their referential situations (Zwaan, 2004). Findings from eyetracking studies have demonstrated that understanding evolves based on the interactive influence between the comprehender, linguistic input, and environmental context.

## 2.1 The Evolution of Simulations

Using what has become known as the 'visual-world' (VW) paradigm, several eyetracking studies have shown that online language processing has an immediate and incremental influence on visual attention (Cooper, 1974; Tanenhaus et al., 1995). In VW studies, participants typically listen to auditorily-presented descriptions while concurrently acting upon or observing an array of visual entities. When people need information about the visual entities in their environment, they simply direct their gaze towards those entities to extract the needed content (location, shape, color, direction, etc.). In the process of looking, observers automatically fixate the most acute portion of the retina, the fovea, onto the the object of interest. This 'foveating' enables the observer to precisely extract the content needed to support ongoing cognition. Thus, the natural eye movements associated with task performance provides a spatially sensitive index of where overt visual attention is deployed. Also, when eye movements are deployed is a precise indicator of the time scale associated with the processes underlying cognition (for a review see Henderson & Ferreira, 2004; Rayner, 1998; Shinoda et al., 2001; Hayhoe et al., 2002). Thus, while listening to descriptions, participants' saccades, fixations, and fixation sequence are measured to identify the real-time influence of linguistic cues on the establishment of reference with visual entities. Although participants' overt responses to linguistic stimuli are typically evaluated to verify understanding, eye tracking has the benefit of capturing an online reflection of the processes underlying language processing before conscious judgments can be made (for a review see Tanenhaus et al., 2004). Thus, the deployment of gaze is a preconscious index of the integration of linguistic and visual representations that reflects how understanding evolves over time as linguistic cues are processed.

Laying the groundwork for contemporary visual-world studies, Cooper (1974) observed the rapid influence of language on comprehenders' visual attention to objects in the environment. During the concurrent processing of spoken linguistic input (stories) and visual context (pictured objects), participants spontaneously directed their gaze to the entity in the visual display whose meaning most closely matched the meaning of the linguistic context. For instance, upon hearing the word 'lion' in a sentence, participants spontaneously shifted their gaze to a picture of a lion. And upon hearing the word 'Africa', participants tended to shift their gaze to semantically related entities (e.g., lion, zebra, or snake). Because these eye movements were not required for successful task completion, it seems that comprehenders spontaneously integrate linguistic and visual content across lexical and semantic representations. Perhaps more surprising is that over half of the saccades made to the target referents were initiated while linguistic input was being processed (i.e., before object names were completely articulated). Thus, after hearing the initial portions of words participants anticipated the object that would be referred to. Cooper posited that comprehenders anticipate what is going to be referred to in ongoing speech and then spontaneously engage the visuomotor system to test their expectations. This spontaneous process appears to show a preconscious influence on behavior.

These studies have also demonstrated that comprehenders exhibit this anticipatory behavior across several dimensions. For example, studies have shown the time course of the activation of perceptual content when perceptual properties are explicitly mentioned or implied. Tanenhaus et al. (1995) demonstrated that comprehenders incrementally establish reference with potential referents based on the overlap between the perceptual features of immediately present objects and linguistic input. Tanenhaus et al. had participants follow instructions like, 'Touch the starred yellow square,' while observing four blocks that had one or more perceptual properties in common (e.g., marking, color, or shape). Participants tended to shift their gaze to the target referent approximately 250 ms after the word that uniquely specified the target. For example, gaze was directed after the word 'starred' if only one starred object was present, and after the word 'square' if only one starred square was present among multiple starred objects. Thus, visual context incrementally influences the resolution of temporary perceptual ambiguities within language, and further demonstrates an immediate interaction between linguistic and visual context such that establishing reference is influenced by the degree of ambiguity present in either.

Providing further demonstrations of the influence of perceptual properties on understanding, Dahan & Tanenhaus (2005) and Huettig et al. (2004) investigated whether the shape structure of pictured objects interacts with the visual representation that is activated from spoken input. Participants observed an array of four pictured objects while hearing individual object names. For example, participants heard the name 'snake' while observing four pictures, one of which was a pictured snake. Critically, one visual competitor was also presented in the display that shared a potential shape structure with the target concept. For example, the target object (e.g., snake) was presented in conjunction with an object that could share the same shape (e.g., length of rope). In this study, these two objects were presented in different shape configurations such that a picture of the target (e.g., a coiled snake) was not visually similar to the competitor (e.g., an elongated piece of rope). Thus, the visual similarity between the visual representations of the referent and the competitor was minimized, such that a non-coiled snake would be paired with a coiled length of rope. It was observed that participants were more likely to fixate on a pictured object that matched a possible structural configuration of the named target object (e.g., a coiled rope would match the potential shape of the elongated snake ) than to those objects that did not even though there was no overlap in the phonetic either of the objects' names. In other words, when participants observed a picture of an elongated snake in conjunction with a coiled length of rope, the length of rope would compete for observer's attention although it was visually dissimilar from the target. In a similar study, Huettig et al. (2004) demonstrated that this structural priming occurs during spoken language at the offset of the object's name, demonstrating that reference is established to the competitor object while linguistic input is processed.

These studies demonstrate that the probability of target fixation reflects the level of congruence between the representation of immediately perceived objects with those representations activated by spoken language. The results indicate a visual-competition effect that can be interpreted in terms of a match between the perceptual representations activated by spoken words and the immediately present objects. Therefore, eye movements appear to be mediated by a re-enactment of experiential traces associated with the various perceptual forms an object can have (Zwaan et al., 2002). Thus, the match between word and object appears to occur at the level of visual features as well as phonemic form. When a linguistic expression refers to an object in a circumscribed visual context it is similar to a natural visual search in which the "top-down" target representation, activated from the spoken input, is mapped onto the "bottom-up" scene representation (e.g., Rao et al., 2002). The results are in line with situated vision accounts, in that overt attention may be re-directed immediately towards at least partially matching objects in order to acquire more information about them. This attentional allocation is driven by the goodness-of-fit between the target referent and the immediately present visual entities. These findings also show that perceptual competitor effects are not limited to 'goal-directed' task demands (Dahan & Tanenhaus, 2004), as perceptual competitor effects were also found with a 'passive' listening task. Huettig et al.'s findings do not speak to whether the visual information in the visual field activates concepts that lead to earlier access of shape information during word comprehension. Therefore, determining the extent to which perceptual property activation originates from the visual field v. linguistic input requires further investigation.

A series of studies by Altmann and colleagues have shown that the understanding of described events involves the anticipation of what environmental entities will be referred to given real-world knowledge (Altmann & Kamide, 1999; Altmann, 2004; Altmann & Kamide, 2004). Altmann & Kamide (1999) had participants listen to sentences such as 'The man will eat the cake' while viewing an array of four entities: two agents (e.g., a man and a woman) and two objects (e.g., a piece of cake and a newspaper). In this example, only one of the objects (cake) satisfied the unique requirements of the verb used in the sentence (eat). As participants listened to the sentences their gaze was directed to the item in the visual scene that satisfied the verb (cake) rather than to those that could not. Given the time scale of the eye movements to the targets, the saccades to the target objects must have been programmed while the verbs were being spoken. Thus, knowledge about events and the objects associated with them were activated almost immediately upon interpreting the disambiguating verb. This interpretation then influenced the deployment of attention accordingly. These findings suggest strongly that background knowledge associated with visual entities influenced comprehenders' expectations of the described events. The timing of the attentional shifts suggest that the simulations evolve as verbal information is interpreted, and reference with environmental entities can be established once there is sufficiently disambiguating information.

The allocation of visual attention to mentioned objects seems tacitly functional. Attending to mentioned objects supports understanding by enabling one to acquire further information about those objects as needed (O'Regan, 1992; Findlay & Gilchrist, 2003; Henderson & Ferreira, 2004). The studies above demonstrate how the allocation of visual attention is influenced during language comprehension. However, some VW studies have also shown that observers direct their attention to the previous locations of objects after the items have been removed from view (Spivey & Geng, 2001; Richardson & Spivey, 2000; Altmann, 2004). For example, Altmann (2004) demonstrated that observers look at the location of a previously seen object upon hearing a description that refers to that object. Thus, while processing a sentence like 'The woman read the newspaper', eye movements were directed to the previous location of the target (newspaper). This effect occurred despite the fact that there was obviously no stimulus present at that location from which to extract information. Another interesting aspect of the findings from this study is that the looks were initiated during the articulation of the verb even before the objects name was mentioned. This demonstration suggests that comprehenders encode the oculomotor coordinates of objects in the environment, and subsequent reference to those objects activates their coordinates and deploys gaze accordingly.

Eye movements have also been shown to reflect the anticipation associated with the simulation of the described state changes (Altmann, 2004; Altmann & Kamide, 2004). For example, Altmann & Kamide (2004) demonstrates that anticipatory gaze shifts reflect when an internal simulation is incongruous with previous experience. In their study, participants observed an array of objects (including among other things a wine bottle and a glass). Then they listened to short passages that described one object (glass) being moved from its original location (on the floor) to a new location (on the table). Subsequent reference to the glass (The woman poured the wine into the glass) caused participants to direct their gaze more frequently to the updated location of the glass (table) rather than to the original location where the object was previously seen. This demonstrates that the oculomotor coordinates encoded during scene observation are not fixed, but can be dynamically updated via language. Thus, comprehension involves a simulation of described events and adjusts the internal coordinates of objects accordingly.

It is important to identify the extent to which components of simulation are actively updated. In a follow-up study, Altmann & Kamide (2004) presented sentences such as, 'The woman will put the glass on the table. Then, she will pick up the wine, and pour it carefully into the glass'. In the visual scene, the winebottle and the glass were located on the floor rather than the table. When the referent was described as having been moved to another location (e.g., 'the glass was moved to the table'), participants tended to look less frequently to the previous location of the pictured glass and more frequently to the linguistically updated location of the glass (i.e., on the table). Thus, without perceiving any visual change, the mental location of the visual entities appear to be updated in accord with its newly described location. In a subsequent study, participants heard sentences such as 'The woman will [clean/break] the glass. Then she will pour the water into the glass', while viewing an array comprised of a woman, a bottle of water, a mug, and a glass. After hearing that the glass was broken, participants tended to shift their gaze immediately to the remaining object that could afford to contain the water (i.e., mug) than when the glass had bean cleaned. Thus, the plausibility of events linguistic descriptions influenced peoples' expectations. These findings suggest that the oculomotor coordinates of visually encoded entities are activated during comprehension and are updated in response to the internal simulation of described events. Therefore, the deployment of attention reflects the mental representation of spatial coordinates rather than the external space itself.

One issue that becomes apparent for using the visual-world paradigm is that the understanding of described events may not be reflected in the types of visual arrays used. Most of the visual-world studies have presented participants with visual context comprised of objects that are functionally, semantically, and spatially incoherent. Investigating the nature of mental simulation may benefit from a better understanding of the influence of varying levels of coherence between visual context and linguistic expressions. For example, if understanding occurs within an embodied comprehender who is necessarily situated in the environment, then the properties of internal representation may be modulated by the comprehenders domain of interpretation and the integratibility of visual and linguistic representations (Richardson & Spivey, 2000). Thus, manipulating visual context may prove to be a useful methodological manipulation to identify the depth of integration between internal representations and the entities in external scenes, and to determine if this integration is influenced by the presence of thematically (in-)coherent objects amid (un-)fitting environmental context.

Comprehending described events appears to involve the activation of the coordinates associated with referential entities. It remains unclear if the activation of spatial representations is dependent on the presence of distinct entities. It could be that observers of the types of scenes used in VW studies develop a strategy of developing a visual-list of the pictured objects. Since it has been shown that the spatial layout of referential entities can affect comprehension (Carlson-Radvansky & Radvansky, 1996), it is important to identify whether or not spatial representations are involved in simulations that involve individual entities in the environment. The simulation of described events may influence the deployment of attention. Such simulations are based on background knowledge about where events occur. Therefore, it is predicted that the construal of described events involves a simulation of object interaction. The site of this interaction should be activated through attention, and then visual attention should be deployed accordingly. These attentional shifts should be reflected in comprehenders' eye movements as they coordinate their internal representations with the environment.

If language processing involves the activation of functional relations of objects, in addition to the activation of object-specific information, then we should see language influence eye movements to particular areas in naturalistic layouts given contextual knowledge. Thematic and spatial coherence between objects in a natural visual scene may involve the anticipation of reference from simulated events that would not emerge with simple object arrays. For instance, linguistic content may activate high-level affordances between cohesively oriented objects whereas discrete object arrays may elicit more low-level property activation (e.g., lexical form, object shape). Many of the picture-priming effects discussed at the beginning of the paper have looked at the constraining effect of context on object properties (e.g., shape: Zwaan et al., 2002; object orientation: Stanfield & Zwaan, 2001). However, it may be useful to also investigate the influence of event knowledge on linguistic understanding given the scene context. In this way, the juxtaposition of scene context with event descriptions may show how language influences the active inspection of natural scenes.

# CHAPTER 3

# PRESENT STUDY

Consider the following sentences:

1. The farmer fed the cow.

2. The farmer milked the cow.

3. The farmer groomed the cow.

Sentences (1)–(3) describe an agent interacting with an object in three different ways. What is implicit with each of these sentences is that each of the verbs are differentially associated with specific spatial regions within the direct object. For example, the verb ‘fed’ in sentence (1) implies that the subject of the sentence (farmer) will interact with the feeding region (head) of the object. Whereas the verb ‘milked’ in sentence (2) implied that the subject will interact with the milking region (udder) of the object. Thus, the verbs used in sentences (1) and (2) implicitly focus on the front and back of the direct object. However, the verb ‘groomed’ in sentence (3) implicates a less specific region of the direct object. It is predicted that comprehenders will activate implicit knowledge about the loci of object interaction, if comprehension does invoke mental representations that are analogs of the sensorimotor representations that accompany the referential events. It is therefore hypothesized that the processing of sentences like ‘The farmer [fed/milked] the cow’ will activate different mental simulations of the referential entities that capture the contextually relevant properties of the situations. Although the differences between the described events are more extensive (semantic, perceptual, and motoric), in this study the focus is on the implicit spatial interaction between entities.

The context-specific spatial activation is predicted to influence the evolution of the mental simulation over time, such that any spatial activation will occur once enough disambiguating linguistic content is processed. This spatial activation should therefore differentially influence comprehenders' allocation of attention over time. Since the linguistic system appears to be biased to establishing reference with objects in the visual world (e.g., Richardson & Spivey, 2000), it is hypothesized that simulated events will automatically activate the oculomotor coordinates associated with the regions of interaction between the environmentally situated objects, and will be accompanied by overt shifts in gaze to that region. Therefore, descriptions in which the locus of the interaction is implicit (e.g., 'The farmer milked the cow'), should influence how comprehenders coordinate their attentional preference in the mental simulation with the environmentally situated referents. This localized activation should, in turn, influence comprehenders' deployment of visual attention and their online visual interaction with the environment. Three experiments were designed to test whether the processing of event descriptions involves the online deployment of visual attention to spatial regions within individual referential entities that are implicitly relevant for the described event. Furthermore, what is novel about this study is that spatial regions within objects, rather than the objects themselves, differentially satisfied the restrictions of the verbs used across conditions.

## 3.1 Experiment 1

This experiment was designed to test whether the comprehension of described events deploys attention to contextually appropriate spatial regions that afford a described event. It is hypothesized that comprehenders perform internal simulations of referential events that implicitly activate context-specific knowledge of event locus within object representations. The language-based object representation should be parallel with the comprehender's visual representations of environmental entities. It is thus expected that this integration of the representations will activate the oculomotor coordinates of the region within the pictured object that affords the described event. The real-world knowledge that is activated during the comprehension of described interactions between objects, should involve the simulation of where the objects will interact. The simulation should then direct attention to the coordinates associated with that region. Thus, the coordination of internal simulations with environmentally-situated referents should be reflected in allocation of gaze to situation-specific regions.

### 3.1.1 Method

**Participants.** Thirty undergraduate students enrolled in introductory psychology courses at Florida State University participated for partial course credit.

**Stimuli.** Sixty-four sentences were constructed: 12 experimental sentence triads and 52 filler sentences. Each experimental sentence set described an agent interacting with an object in three different ways (e.g., 'The farmer [fed/milked/groomed] the cow'). In each experimental triad the verb was the only term manipulated between sentences. Two of the verbs were selected because they implied an asymmetrical spatial interaction with the direct object. For example, in the sentences above the verb 'fed' implies a farmer-head interaction and 'milked' implies a farmer-udder interaction. However, in addition to these two region-specific verbs, a third neutral verb (e.g., 'groomed') was included that did not specify a locus of interaction. In addition to the experimental set, 52 additional sentence-picture pairs were used as fillers. These items included the same type of objects, but the sentences included verbs that were less-restrictive in the spatial regions within the objects that they referred to (e.g., the boy dropped the ice cream). All 12 critical items, and 20 of the filler items, were paired with a picture of the direct object. However, 32 of the filler sentences were paired with a picture of an unmentioned object. Thus, there were an equal number of sentences that mentioned the pictured object and sentences that did not.

Twelve experimental pictures (see Figure 3.1) were each paired with three sentence conditions corresponding to specific spatial activation as in sentences (1) & (2), or more generalized spatial activation as in sentence (3).

1. The farmer fed the cow.

2. The farmer milked the cow.

3. The farmer groomed the cow.

The 64 pictured objects were comprised of grayscale line-drawings and photographs. All were scaled to occupy a region of approximately 20 cm. Twelve pictures were paired with the experimental sentences, and the remaining 52 pictures were paired with the filler sentences. Each experimental picture displayed a profile of the objects as they would be seen from the

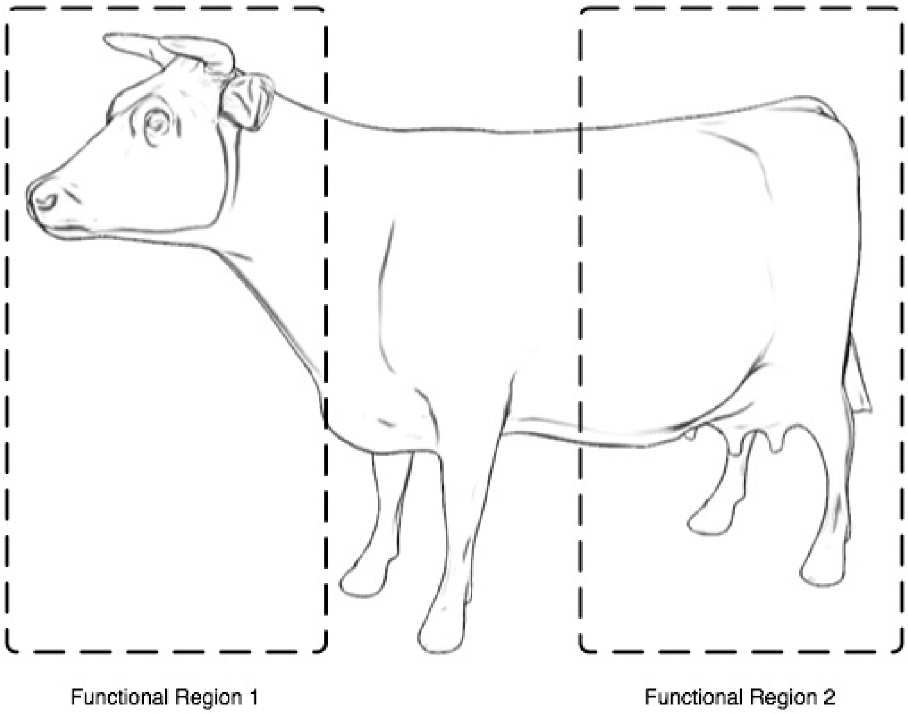


Figure 3.1: Example stimulus with circumscribed regions of functional interaction.

side. This side-view presentation enabled each item to have its distinct verb-restricted regions at opposite ends (e.g., a cow with its head to one side of the screen and its hind-region to the opposite side). To counterbalance across conditions the orientation of each experimental picture (i.e., left- v. right-facing) was flipped along its vertical axis. Counterbalancing the horizontal orientation ensured that any attentional preference to more visually salient regions of the object would be eliminated.

Six stimulus lists were created to factorially counterbalance items and conditions. This produced a 3 (Sentence: Front-biasing/Neutral/Back-biasing) X 2 (Orientation: Left-facing/Right-facing) design. Sentence was a within-participants and within-items factor. Although orthogonal to the manipulation of interest, the orientation of the pictured objects was manipulated to properly counterbalance the spatial orientation of objects. Thus, participants across conditions viewed the same object flipped along its vertical axis. Each participant saw only one of six possible versions of each sentence-picture pairing.

Stimuli were presented in a randomized order for each participant. Each participant heard only one version of each sentence set. The sentences were created with a computer-generated voice. A U.S. English male "David" speaking voice from Cepstral (Cepstral Inc.http://www.cepstral.com) was used at a sampling frequency of 44.1 kHz. The prosody of each sentence was normal, however the speech was slowed slightly to maximize clarity (the stage durations averaged across sentences are presented in Table 3.1).

Table 3.1: Mean Offset Latencies (Standard Deviations) in ms for each stage of the spoken sentences in Experiments 1–3.

| 'The farmer | fed | the cow.' |
|---|---|---|
| 595 (115) | 950 (164) | 1556 (190) |

**Apparatus and Procedure.** Stimulus presentation was controlled by E-Prime (Psychology Software Tools, http://www.pstnet.com). Pictorial stimuli were presented on a 21" flat-screen CRT operating at refresh frequency of 85 Hz. Auditory stimuli were presented over loudspeakers. Each participant was able to adjust the volume of the auditory stimuli to a comfortable setting.

Eye movements were recorded using an Applied Science Laboratories, Inc. (http://www.a-s-l.com) EyeTrac 6000 head-mounted infrared eye tracker with magnetic head-tracking. The eyetracker sampled participants gaze at a frequency of 60 Hz. Although not actively controlled, average viewing distance was approximately 65 cm. The magnetic head-tracking unit was responsible for monitoring and correcting the participants' gaze in relation to the their posture in real-time. Initial setup involved an experimenter-paced 9-point calibration and general instruction about the task before participants began the study.

During each trial, a fixation point was presented in the center of the screen. Participants were instructed to look directly at the fixation point, and then press the spacebar to begin the trial. Upon pressing the spacebar, the fixation point changed to red for 250 ms, and then a pictured object was presented in the center of the screen. Then, 250 ms after picture onset, a recorded sentence was presented auditorally over speakers while participants continued to view the pictured object. Participants were instructed to pay attention to both the picture and the sentence in each trial. They were also told that in some trials the sentences would describe something that could happen with the pictured object, and others would not. The picture remained on the screen for 4 s, and then a text-box appeared instructing the participant to enter the name of the depicted object. Participants were encouraged to respond as accurately as possible as accuracy of naming would be measured. The object names entered by the participants were recorded to ensure that each object was attended to and correctly identified. The entire procedure required approximately 45 minutes.

### 3.1.2 Results

Two spatial regions of interest (matching and mismatching; see Figure 3.1) were circumscribed for all pictures. Because the critical items were always presented in profile, these regions always corresponded to the front and back of objects. The dependent measure was the mean proportion of trials that included fixations to the matching v. mismatching spatial regions. For example, after hearing 'the farmer fed the cow', looks towards the cow's head were compared with looks towards the cow's udder. Thus, each item served as its own control (looks towards the mismatching region provided the baseline against which to compare looks towards the matching region). This design ensured that any potential effects caused by differences in visual salience amongst the regions were eliminated. Furthermore, the two spatial regions were segregated by a middle region of approximately 5 cm to exclude ambiguous fixations.

The proportions of looks to both spatial regions were analyzed across time. Six temporal regions were identified for each sentence. The temporal regions captured the duration of the 1) the first noun phrase, 2) the verb, and 3) the second noun phrase. These three temporal regions were defined by the auditory time course for each experimental sentence. The average durations for each of the temporal regions are reported in Table 3.1. Three temporal regions corresponded to the stage of each sentence such that the first epoch included the time from sentence onset to the offset of the subject (i.e., 'The farmer'). The second epoch included the duration of the critical verb ('fed'). The third epoch included the onset of the second noun phrase ('the') until the offset of the direct object ('cow'). Additionally, three 300 ms temporal regions were defined post-sentence to capture any long-term effects. Epochs 4–6 began at sentence offset and therefore were also defined by each experimental sentence. Together, epochs 4–6 captured the fixations that occurred within 900 ms period following each sentence. All fixations outside either the critical spatial or temporal regions were omitted from analysis.

The proportion of trials with fixations to the matching v. mismatching regions were submitted to 2 (spatial region) X 6 (temporal stage) analyses of variance (ANOVAs) with repeated measurement on region and stage. All looks outside of the predefined spatial and temporal regions were omitted from the analyses. Because the verbs used in the neutral condition were not designed to be paired with the predefined matching v. mismatching regions, they were omitted from these analyses. The mean proportion of trials with fixations to the critical regions and synchronized with speech input are displayed in Figure 3.2.

Contrary to the predictions, there was no interaction between region and stage ($F < 1$). Nor was there a main effect of match ($F < 1$). Thus, linguistic content did not influence participants' observation of regions within individual objects. However, there was a main effect of stage in which the proportion of trials with looks to the critical regions changed as the sentences evolved ($F(5,175) = 10.09$, $p < .001$, $MSe = .22$). The proportion of looks to the critical region was maximal during the processing of the second noun phrase ('the cow'). This effect is unsurprising given the design of the task where participants begin each trial by looking at the center of the screen and then look farther from the center over time. This pattern holds for both the matching and mismatching spatial regions. [1] The finding of a significant result for stage suggests that the lack of difference between the match condition and mismatch condition is not simply the result of insufficient power or sensitivity of the dependent measure. Potential explanations for this finding will be discussed in Section 4.1.

[1]The data from Experiments 1–3 were also submitted to additional analyses to get a more complete picture of comprehenders' fixation behavior during comprehension. These analyses yielded the same pattern of results, so for the sake of exposition they will be described here briefly. In addition to the mean proportion of trials with looks (e.g., Altmann, 2004) the mean proportion of fixations to the matching region were compared with those fixations to the mismatching region for each trial. Given the size of the temporal bins, the most participants only fixated once or twice during each bin. Thus, strictly speaking the fixation data were almost identical to the trial-based counts presented here. Also, the size of the spatial bins were manipulated to determine if an effect of region required a more fine-grained spatial analysis. However, the size of the spatial bins was ultimately selected because it seemed to be a suitable compromise for capturing sufficient fixation data while omitting ambiguous looks to the front v. rear of objects. Although these alternative analyses did not yield results different from those presented in the main text, it is worth noting that pictorial and verbal stimuli with more consistent temporal and spatial overlap than those used in the present study may make such analyses preferable.

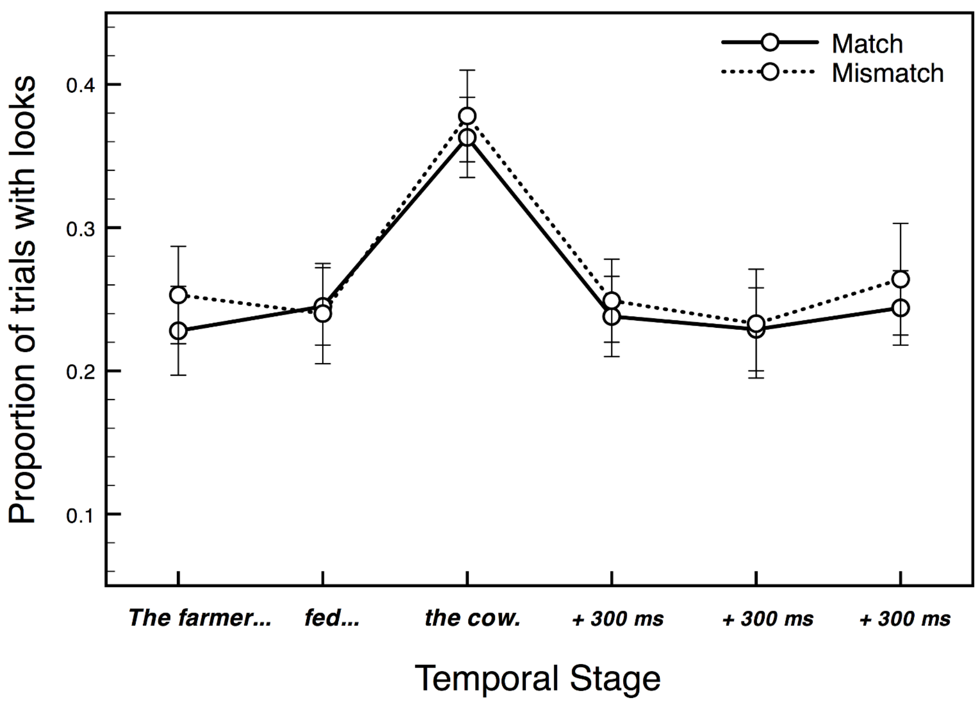


Figure 3.2: Mean Proportion of looks to circumscribed regions in Experiment 1

## 3.2 Experiment 2

A few visual-world studies have demonstrated that the allocation of gaze during linguistic processing is not dependent on the presence of visual entities (Richardson & Spivey, 2000; Spivey & Geng, 2001; Altmann, 2004). For example, Altmann (2004) demonstrated this with an altered version of the visual-world paradigm called the 'blank-world' paradigm. In this study, participants listened to sentences like 'The man will eat the cake' after the visual arrays were removed from the display. The stimuli were just like those used in Altmann & Kamide (1999) with visual arrays comprised of two agents and two objects, and each of the objects satisfying the requirements for one of the verbs used in sentences. When the verb ('eat') and object name ('cake') were mentioned eye gaze was consistently directed to the location where the matching referent was previously presented. Apparently, the triggering of eye movements during linguistic expressions is not contingent on the visual entities' presence. Rather, observers seem to encode the oculomotor coordinates of objects and processing an expression that refers to such objects activates those internal coordinates. Thus, the deployment of visual attention during language processing is not contingent on the locations of visible objects, but rather on an observers' internal representation of the entities in the environment. Moreover, the results indicate that participants' eye movements were launched during the processing of the verb which is similar to the findings of Altmann & Kamide (1999), suggesting that the time course of visuolinguistic integration is the same whether or not the visual world is visible.

Experiment 2 was designed to test if the processing of described events influences comprehenders' visual processing once the visual inspection of an object has been completed. The primary rationale for performing this experiment was to determine to what extent the influence of mental simulations on one's visual interaction with the world is dependent on object presence. The allocation of gaze to the context-specfic regions is predicted to be independent of object presence given previous findings (Richardson & Spivey, 2000; Spivey & Geng, 2001; Altmann, 2004). The failure to find an effect of attentional preference in Experiment 1 raises the possibility that any attentional preference for regions within individual objects requires a more complete encoding of visual entities before linguistic processing (although there is evidence suggesting that longer exposure times are unnecessary to elicit systematic eye movements to referents during concurrent linguistic and visual processing, e.g., Dahan & Tanenhaus, 2004). Although it is unclear if observers automatically encode the oculomotor coordinates of functional regions during early object identification, if observers encode the coordinates of regions within objects with longer observation, then subsequent activation of those coordinates should deploy attention accordingly. Experiment 2 tested whether comprehenders direct their gaze to previously encoded regions within objects during linguistic processing.

### 3.2.1 Method.

**Participants.** Thirty-six undergraduate students were drawn from the same subject pool as the first experiment.

**Stimuli.** The stimuli used in Experiment 1 were unchanged for Experiment 2.

**Design and Procedure.** The difference between Experiments 1 and 2 was the order of stimulus presentation. During each trial a pictured object appeared for 3 s. Once this picture was removed from the display, a gray mask was presented on the display to mask any visual afterimage. After a 1 s interval, each sentence was presented auditorily. After each sentence, participants were prompted to enter the name of the previously seen object. The entire procedure required approximately 45 minutes.

### 3.2.2 Results.

The analyses performed in Experiment 1 were repeated here, but the comparisons of interest include the eye fixations that occurred after the pictured object was removed from view. Thus, the eye movements of interest are those that occurred while the participants observed a blank screen. The circumscribed spatial regions were identical to those used in Experiment 1, except that they refer to the previously seen regions. For the sake of exposition, the regions of interest for the previously viewed objects will simply be referred to as regions of interest.

The predictions put forth in Experiment 1 were the same for Experiment 2. Thus, upon listening to each sentence, participants are expected to be more likely to fixate the matching spatial region v. the mismatching region. This effect would suggest that the coordinates of functional regions are encoded during longer observation, and subsequent mental simulations are sufficient to deploy gaze to those coordinates even when the object is absent from view.

Also as predicted in Experiment 1, the deployment of attention is predicted to occur once the simulation is sufficiently constrained by linguistic context (i.e., during the verb).

As in Experiment 1, contrary to the predictions, there was no interaction between region and stage ($F < 1$) or a main effect of match ($F < 1$). However, there was a main effect of stage in which the proportion of trials with looks to the critical regions changed as the sentences evolved ($F(5,175) = 4.57$, $p < .001$, $MSe = .17$). The proportion of looks to the critical region was maximal during the processing of the first noun phrase ('The farmer'). However, there was no effect during or after the mention of the critical verb. This finding will be discussed in Section 4.1.

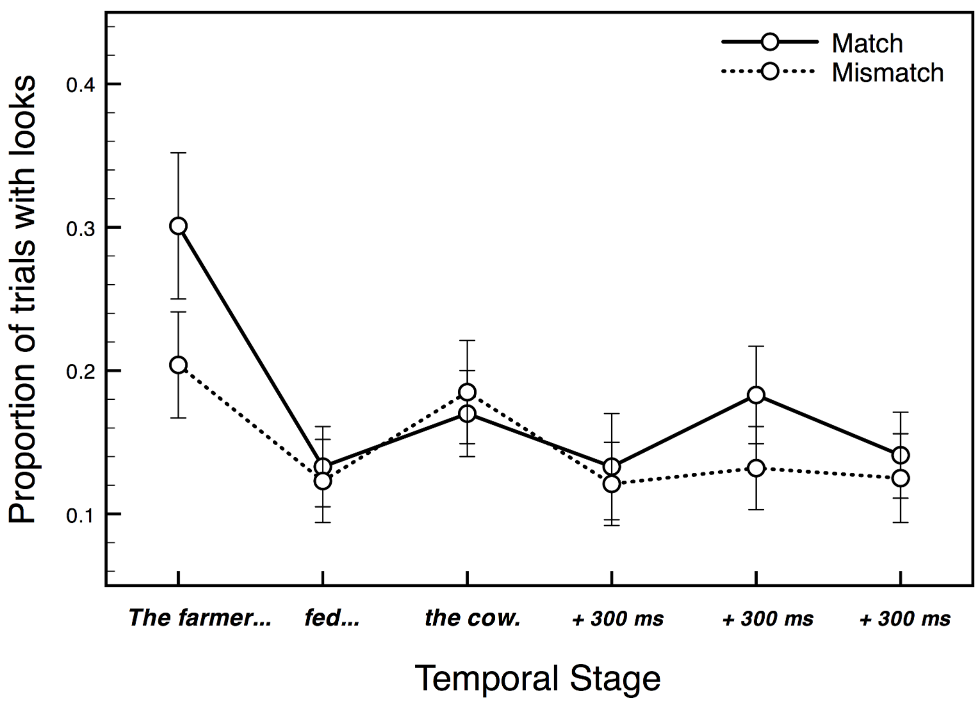


Figure 3.3: Mean Proportion of looks to circumscribed regions in Experiment 2

## 3.3 Experiment 3

One of the main questions regarding the effects of language-guided attention is whether these eye movements play a functional role in comprehension or whether they are epiphenomenal. Experiments 1 and 2 did not demonstrate that online language processing directs attention to context-specific spatial regions for concurrently-viewed or previously-viewed objects. It is possible that the individual objects used in Experiments 1 and 2 enable observers to satisfy the requirements of the task without an exhaustive visual search. Thus, the function of the inspection of visual entities would become less relevant than it would otherwise be and any influence of mental simulation on eye movements would become less pronounced. The findings from Experiments 1 and 2 suggest that the influence of language on eye movements is not simply a matter of conscious allocation of attention. Experiment 3 was designed to test whether language processing directs attention to context-specific regions when objects are noncanonically oriented. If objects in the environment are situated in manner that makes them more difficult to identify and incongruous with the described events, the mental simulation may play a more functional role in how one interacts with the world.

### 3.3.1 Method.

**Participants.** Thirty-six undergraduate students were drawn from the same subject pool as the first experiments.

**Stimuli.** The same stimuli used in Experiments 1 and 2 were used for Experiment 3. Although the sentences were unchanged, the experimental pictures were flipped along their horizontal axis so they appeared upside-down. Twenty filler pictures were also flipped upside-down so that half of the items appeared upside-down and the other half appeared right-side-up.

**Design and Procedure.** The same factorial structure used to counterbalance items and conditions in the first two experiments was used in Experiment 3.

As in Experiment 1, participants were instructed to look at the fixation point, press the spacebar to begin each trial, and then listen to each sentence while attending to the accompanying pictured object. After each trial participants were prompted to enter the name of the depicted object.

The critical difference from Experiments 1 and 2 is that the vertical orientation of the visual stimuli in the experimental trials were manipulated. In the first two experiments, objects were depicted in one of two orientations along the vertical axis (left- v. right-facing). Thus, the objects in these experiments were always canonically situated (i.e., right-side-up). However, in this experiment, the critical items were flipped upside down and therefore not situated canonically in the environment.

### 3.3.2 Results.

The analyses used in Experiments 1 and 2 have been repeated for Experiment 3.

As in earlier experiments, contrary to the predictions, there was no interaction between region and stage ($F < 1$). Nor was there a main effect of match ($F < 1$). However, there was a main effect of stage in which the proportion of trials with looks to the critical regions changed as the sentences evolved ($F(5,175) = 3.0$, $p < .013$, $MSe = .08$). Thus, linguistic content did not influence participants' observation of individual objects even when the objects were in an atypical orientation. This finding will be discussed in Section 4.1.

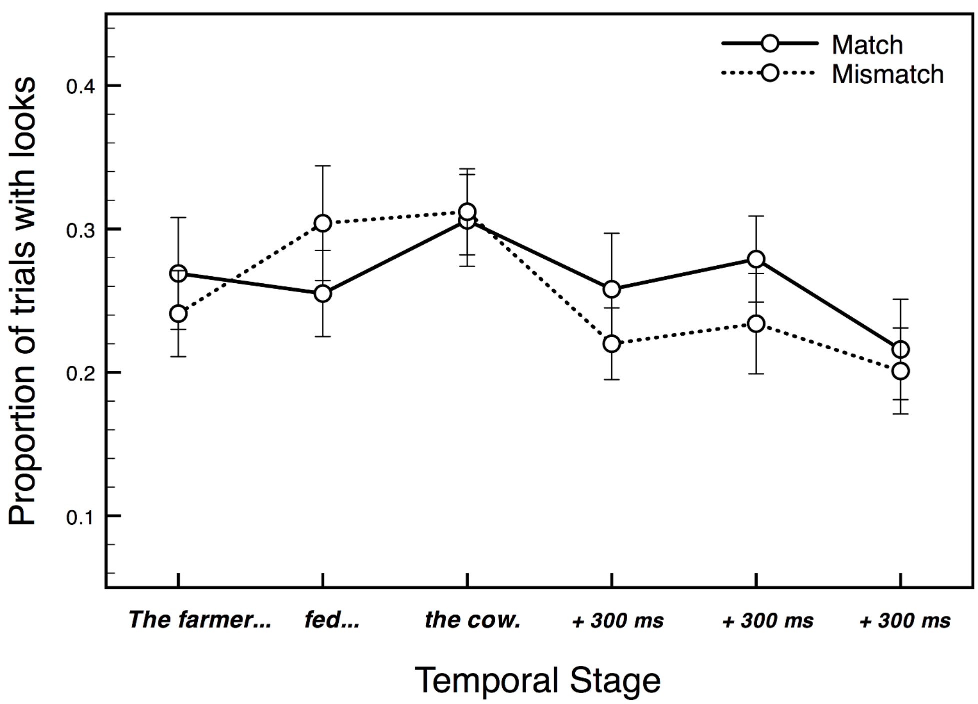


Figure 3.4: Mean Proportion of looks to circumscribed regions in Experiment 3

# CHAPTER 4

# CONCLUSION

## 4.1 General Discussion

This study was designed to address two primary objectives: the degree to which comprehension involves mental simulations and the time course of the processes involved in such simulations. I tested whether language comprehension involves the mental simulation of context-specific regional activation within objects. This idea was based on recent studies that have shown that language comprehension involves the mental simulation of perceptual and motoric activation. The idea was also influenced by recent 'visual-world' studies that have shown that the deployment of visual attention reflects the time course of the processes involved in mental simulation. This study is novel in that it investigated the influence of online mental simulation on visual attention at a more fine-grained inner-object level of analysis rather than the inter-object level of analysis that is commonly used in eyetracking studies. Investigating the specificity of this simulation process was motivated by studies that have demonstrated the activation of object-specific perceptual content during comprehension (e.g., Zwaan et al., 2002; Stanfield & Zwaan, 2001).

It was hypothesized that the processing of described events like 'The farmer fed the cow' would activate a mental simulation of the locus of interaction between the described agent and object. This spatial activation was expected to influence how comprehenders establish reference with objects in the environment. However, counter to the predictions, Experiments 1–3 did not demonstrate that comprehenders direct attention to context-appropriate v. inappropriate regions within objects. Upon listening to 'The farmer fed the cow' people were equally likely to attend the context-appropriate head region v. the context-inappropriate hind region. The findings suggest that comprehenders do not necessarily activate context-specific spatial representations within objects during online language processing. Although this does not support the hypothesis that comprehenders anticipate and establish reference on such a fine-grained level when integrating visual and linguistic content, there are potential factors which may be responsible for mediating and moderating this effect. There are several potential explanations for why this effect did not emerge which will be discussed below.

First, although context-specific spatial activation may occur during mental simulation, visual attention may not be sensitive to this activation if the task does not demand it. Although this possibility cannot be ruled out, many of the VW studies performed by Tanenhaus and colleagues have found attentional effects when participants were instructed to interact with objects (e.g., 'pick up' the object). It is unclear whether such effects would persist without the participants' active engagement with the objects. However, in recent work, Altmann and colleagues have found effects when participants were simply asked to observe visual scenes while listening to descriptions about the objects in the scenes. Although such observation is arguably not passive (O'Regan & Noë, 2001; Findlay & Gilchrist, 2003), it does provide a stronger demonstration that routine comprehension involves the automatic activation of mental simulations. Also, such effects have been demonstrated when stimuli are clearly removed from the participants' view (Richardson & Spivey, 2000; Altmann, 2004). These findings suggest that the allocation of visual attention is robust and not dependent on task demands.

Second, a critical aspect of the study's design is that each sentence described an event that implicitly referred to a region *within* an object. It is possible that regions of objects must be named explicitly in order for comprehenders to establish reference with them. Several VW studies have used explicit mention to investigate the online processing of linguistic cues to understand the incremental nature of phoneme processing and its integration with physical properties of objects in the visual scene (e.g., Tanenhaus et al., 1995). However, Cooper (1974) found evidence not only for language-directed attention to mentioned objects (e.g., looks to a pictured lion upon hearing 'lion') but also to objects that were semantically associated with object names (e.g., looks to a snake and lion upon hearing 'Africa'). More recent work has also demonstrated that comprehenders establish anticipatory reference with objects based on the matching affordances between a described event and object. For example, while viewing a scene with a piece of cake and a newspaper, comprehenders direct their gaze to the object that was congruent with the event described (e.g., 'eat'–cake; 'read'–newspaper) before the actual target object was mentioned. This evidence suggests that implicit reference within a linguistic expression is sufficient to activate individual objects. Despite this, the findings from the present study suggest that implicit reference to regions within objects is insufficient to reliably influence visual attention.

Third, it is also possible that mental representations do not invoke knowledge about regions within objects unless there is an apparent working relationship between visual entities. Several studies have shown that linguistic representations are influenced by the relative spatial positions of objects (for a review see Coventry & Garrod, 2004). For example, in a study by Carlson-Radvansky & Radvansky (1996), the orientation of objects in a scene influenced people to generate descriptions from different perspectives. When the objects were in a functional orientation (e.g., a mailman facing a mailbox) participants adopted an object-centered perspective ('the mailman is in front of the mailbox'), however when they were in a nonfunctional orientation (e.g., a mailman facing away from mailbox) participants adopted a viewer-centered, or deictic, perspective ('the mailman is to the left of the mailbox'). Thus, the spatial relationship between functionally associated objects can influence one's processing of relations and one's frame of reference when describing such relations. However, this effect did not occur with objects that were not functionally associated. This study demonstrates that linguistic representations are influenced by the relative orientation of visual entities, but only when the comprehender can link the object based on an understanding of their working relationship. It remains unknown if this effect occurs automatically during online linguistic processing. It is also unknown whether the orientation of individual objects can influence comprehension.

Fourth, and perhaps most relevant, is the nature of the language itself. A recent study by (Altmann & Kamide, submitted) demonstrates that tense influences comprehension and the establishment of reference. In this study, participants viewed scenes with two objects (a full glass of beer and an empty wine glass) while listening to sentences such as 'The man [will drink/has drunk] the wine'. They demonstrated that looks to the full glass of beer were more likely upon hearing 'will drink' v. 'has drunk', whereas looks to the empty wineglass were more likely upon hearing 'has drunk' v. 'will drink'. Thus, linguistic tense influences one's understanding and expectations of events. The present study relied exclusively on past tense descriptions. Thus, all critical trials described completed actions. It would be interesting to create a future tense version of the present study to determine if comprehenders do allocate attention to target regions when an event is upcoming v. completed. If so, linguistic perspective would obviously be a relevant moderator in comprehenders' ability to establish reference with regions of objects. Given that perspective could lead to distinct representations it may also influence task performance and may yield different levels of performance when tasks demand either action planning for future events or the retrieval of memories from past events.

Each of the four issues raised above are likely to play an influential role in comprehension. First, it seems reasonable to assume that comprehenders would establish reference with environmental objects differently based on task demands. However, some evidence suggests that these effects are automatic and are not disrupted when objects are obviously removed from participants' view (Richardson & Spivey, 2000; Altmann, 2004). Therefore, it seems unlikely that task demands would have much control over language comprehension. Second, it is of course possible that comprehenders do not simulate context-specific regions within objects without processing descriptions with explicit mention of those regions. However, several studies have shown that comprehenders do establish reference based on implicit semantic associations (Cooper, 1974), perceptual overlap (Dahan & Tanenhaus, 2005), and understanding of events (Altmann & Kamide, 1999). Third, the role of environmental context on understanding may require more than individual objects. As Carlson-Radvansky & Radvansky (1996) demonstrated, even multiple objects must be functionally related in order for their spatial relationship to influence understanding. Future work must investigate how comprehenders establish reference differentially with individual v. multiple visual entities. Fourth, the nature of linguistic input (e.g., tense) may influence how comprehension evolves by activating distinct mental representations. It seems quite likely that comprehenders' would generate different expectations if an event were described as being completed v. upcoming. It would be interesting to investigate how tense would affect the understanding of events like those used in the present study.

## 4.2 Conclusion

This study was designed to test whether comprehension involves the simulation of context-specific regions within objects. This hypothesis was based on recent evidence that demonstrates that people use experiential knowledge to make routine inferences about the nature of situations when processing sentences. Such findings provide strong support for the idea that comprehenders mentally simulate described events, and presumably these elaborative inferences fall out of this process. However, the findings from the present study suggest that comprehenders do not automatically activate context-specific spatial representations during online language processing. In other words, the cognitive system does not necessarily deploy a shift of visual attention to support understanding.

Although converging evidence from a diverse set of paradigms suggests that sensorimotor simulations routinely accompany comprehension, the functional role of such sensorimotor simulations remains an open question. The specificity of meaning available to a comprehender may depend on the degree of correspondence between linguistic and visual information. For example, it is likely that linguistic cues such as tense play a critical role in how the cognitive system generates simulations, anticipates events, and directs attention. It may also be the case that the number, type, and placement of objects in the environment enable comprehenders to generate expectations, whereas individual objects do not. Additionally, task-demands may differentially activate specific types of sensorimotor content when generating such simulations. Future investigations are necessary to establish the extent to which conceptual processing is dependent on such factors, and when comprehenders overtly investigate the correspondence between linguistic expressions and referential objects.

Understanding the format of content involved in mental representation is necessary for theory on mental simulation to evolve. Critical to this mission is investigating how conceptual and sensorimotor representations interactively support comprehension. Future studies are planned to further investigate how linguistic tense influences the construction of unique conceptual representations over time. It will be interesting to see if the allocation of attention in this future study is indicative of a conceptual system that establishes reference with context-specific regions for descriptions about upcoming events. Such findings would not only indicate that sensorimotor simulations do involve precise spatial activation, but that establishing such specific reference only occurs when the cognitive system deems it relevant to do so.

# APPENDIX

The sentential stimuli used in Experiments 1–3. The three versions of each item are presented on the same line.

1. The photographer [focused/loaded/stored] the camera.
2. The father [jumpstarted/packed/washed] the car.
3. The farmer [fed/milked/groomed] the cow.
4. The cook [filled/grabbed/washed] the pan.
5. The soldier [cleaned/loaded/fired] the pistol.
6. The nannie [loaded/pushed/cleaned] the stroller.
7. The waiter [poured/grabbed/brewed] the tea.
8. The custodian [cleaned/repaired/installed] the toilet.
9. The mover [connected/loaded/painted] the trailer.
10. The mechanic [tuned/loaded/inspected] the truck.
11. The musician [muted/tooted/polished] the trumpet.
12. The gardener [filled/hoisted/fixed] the wheelbarrow

# REFERENCES


Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: the 'blank screen paradigm'. *Cognition*, *93*(2), B79-87. 2.1, 1, 3.2, 4.1

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-64. 2.1, 3.2, 4.1

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), (Vols. The interface of language, vision, and action: Eye movements and the visual world). New York: Psychology Press. 1, 2.1

Altmann, G. T. M., & Kamide, Y. (submitted). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Cognition*. 4.1

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577-609; discussion 610-60. 1, 2

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philos Trans R Soc Lond B Biol Sci*, *358*(1435), 1177-87. 2

Carlson, L. A., & Kenny, R. (2006). Interpreting spatial terms involves simulating interactions. *Psychonomic Bulletin and Review*. 2

Carlson-Radvansky, L. A., & Radvansky, G. A. (1996). The influence of functional relations on spatial term selection. *Psychological Science*, *7*, 56-60. 2.1, 4.1

Cooper, R. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84-107. 1, 2.1, 4.1

Coventry, K. R., & Garrod, S. C. (2004). *Seeing, saying, and acting: The psychological semantics of spatial prepositions*. New York: Psychology Press. 4.1

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *J Exp Psychol Learn Mem Cogn*, *30*(2), 498-513. 2.1, 3.2

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: conceptually mediated eye movements during spoken-word recognition. *Psychon Bull Rev*, *12*(3), 453-9. 2.1, 4.1

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing.* Oxford University Press. 2.1, 4.1

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*(1), 1-19. 1

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, *9*(3), 558-65. 1

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585-612. 1

Harnad, S. (1990). The symbol grounding problem. *Phys. D*, *42*(1-3), 335–346. 1

Hayhoe, M., Aivar, P., Shrivastavah, A., & Mruczek, R. (2002). Visual short-term memory and motor planning. *Prog Brain Res*, *140*(0079-6123 (Print)), 349-63. 2.1

Henderson, J. M., & Ferreira, F. (2004). In the interface of language, vision, and action: Eye movements and the visual world. In I. J. M. Henderson & F. F. (Eds.) (Eds.), (p. 1-58). New York: Psychology Press. 1, 2.1

Huettig, F., Gaskell, M. G., & Quinlan, P. T. (2004). How speech processing affects our attention to visually similar objects: Shape competitor effects and the visual world paradigm. In *Proceeedings of the 26th annual meeting of the cognitive science society,* (p. 606-612). 2.1

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., et al. (2005, Jan). Perception of motion affects language processing. *Cognition*, *94*(3), B79–B89. 1

Loftus, E., & Palmer, J. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 585-589. 2

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: the world as an outside memory. *Canadian Journal of Psychology*, *46*(3), 461-88. 1, 2.1

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939-73; discussion 973-1031. 4.1

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Res*, *42*(11), 1447-63. 2.1

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422. 2.1

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and hollywood squares: looking at things that aren't there anymore. *Cognition*, *76*(3), 269-95. 2.1, 3, 3.2, 4.1

Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Res*, *41*(25-26), 3535-45. 2.1

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychol Res*, *65*(4), 235-41. 2.1, 3.2

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156. 1, 4.1

Tanenhaus, M. K., Chambers, C. G., & E, H. J. (2004). Referential domains in spoken language comprehension: Using eye movements to bridge the product and action traditions. In J. M. Henderson & F. Ferreira (Eds.), (Vols. The interface of language, vision, and action: Eye movements and the visual world., p. 279-317). New York: Psychology Press. 2.1

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-4. 1, 2.1, 4.1

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*(4), 625-36. 1

Yaxley, R. H., & Zwaan, R. A. (in press). Simulating visibility during language comprehension. *Cognition*. 1

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation: Advance in Research and Theory*, *44*, 35–62. 1, 2, 2

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168-71. 1, 2, 2.1, 4.1

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin and Review*, *10*(4), 954-8. 1, 2

# BIOGRAPHICAL SKETCH

**Richard H. Yaxley**

Richard H. Yaxley was born in October of 1976 in Gainesville, Florida. Richard graduated magna cum laude from Embry-Riddle Aeronautical University in 1999, receiving his Bachelor of Science degree in Aerospace Studies. He entered the graduate program in Cognitive Psychology at Florida State University in 2000 to work with Professor Rolf A. Zwaan. He received his Masters of Science degree in this program in 2003. Richard's research focuses on the mental representation of knowledge, and how cognitive mechanisms recruit sensorimotor content to support understanding.