Agents and Affordances: Listeners Look for What They Don't Hear

Caitlin M. Fausey (cmfausey@psych.stanford.edu)
Department of Psychology, Stanford University
Stanford, CA 94305 USA

Teenie Matlock (tmatlock@ucmerced.edu)
Cognitive Science Program, University of California, Merced
Merced, CA 95344 USA

Daniel C. Richardson (dcr@ucsc.edu)
Department of Psychology, University of California, Santa Cruz
Santa Cruz, CA 95064 USA

Abstract

How do implicit aspects of language guide overt perceptual behavior? In this eyetracking study, we examined whether different ways of describing objects and actions would influence the visual processing of objects with affordances. Specifically, we were interested in the effect of different information about the agent of an action. English-speaking adults viewed objects with interactive regions, such as handles, knobs or buttons. Participants viewed each object after listening to a sentence with or without information about an agent. Participants were faster to fixate the interactive region of objects after hearing non-agentive language than after hearing agentive language, as if they were searching to fill an "agent information gap". These results may inform theories about how global knowledge and local linguistic information mutually determine visual inspection of objects.

Keywords: Affordances; Language-mediated eye movements

Introduction

Much of our everyday understanding of physical objects is grounded in affordances. This includes tacit knowledge about how objects are canonically oriented, what they are used for and, critically, how we interact with them. We know, for instance, that pitchers have handles for pouring, cars have steering wheels for driving and guns have triggers for shooting. The current study examines object affordances at the interface of language and visual processing. Do different linguistic environments change how people visually inspect objects that afford human action? Specifically, how might language that differentially codes for agency guide attention to interactive regions of these objects?

The notion that visual, motor and linguistic representations are tightly linked has received empirical support in recent years (e.g., Barsalou, 1999; Glenberg, 1997; Pecher & Zwaan, 2005). For example, Tucker and Ellis (1998) found that people were faster to judge whether a cup was right side up or upside down when the cup handle was on the same side of the screen as the hand with which they made their response than when the handle was on the opposite side of the response hand. Glenberg and colleagues have observed similar "action compatibility effects" in language comprehension and judgment tasks. For example, when discriminating sensible from nonsensical sentences, participants answered fastest when the location of the response was consistent with the movement described by the sensible sentence, as in pressing a button close to the body after reading "*open the drawer*" (Glenberg & Kaschak, 2002). In a part-judgment task, participants were faster to verify parts toward the upper half of objects when they made responses requiring upward movement, and lower parts with downward movement (Borghi, Glenberg & Kaschak, 2004).

Evidence for tight links between semantic and motor representations of object affordances has been found when movements themselves are the dependent measure. Creem and Proffitt (2001) found different grasping behavior when participants either did or did not concurrently perform a semantic task while grasping. Without an additional task (or with an unrelated spatial task), participants grasped objects such as combs, spatulas and paintbrushes by their handles. When completing a concurrent semantic task, this normal grasping behavior was disrupted. This effect suggests that normal object-directed movement relies on semantic knowledge about object affordances.

Even when no overt response is required of experimental participants, representations of object affordances may still be active. One source of evidence in support of this claim is the finding that neural circuits that are activated during grasping are also activated when people simply view manipulable objects (Chao & Martin, 2000). Additional evidence that suggests an automatic activation of knowledge about object affordances comes from eyetracking studies.

Affordances and eyetracking. Eyetracking provides one measure of how people integrate background knowledge, language and visual information in real time. Researchers have studied the interaction of eye movements and linguistic processing in various ways: Many studies have examined the contribution of eye movements to resolving ambiguities in sentence understanding (e.g., Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995) while others have reversed the question and examined the influence of language itself on visual processing (e.g., Richardson & Matlock, 2007).

In general, eye movement data suggest that listeners are sensitive to the semantics of verbs (and co-occurrences with nouns) when integrating linguistic, visual and motor information. For example, in a study by Kamide, Altmann and Haywood (2003), participants made more anticipatory eye movements to an image of butter after hearing the verb *spread* than after hearing the verb *slide*. Using a task that directed participants to move objects, Chambers, Magnuson and Tanenhaus (2004) found that changing the non-linguistic context – whether objects were picked up by hand or by hook – influenced sentence comprehension. Sentences that had been ambiguous in the by-hand context became unambiguous in the by-hook context because only one object could be picked up using a hook. Participants' eye movements reflected the lessened ambiguity in the hook context as soon as 150 ms after mention of the potential referent. Thus, knowledge about the affordances of "picking up" was integrated with visual and linguistic information in this task.

Where people look for information in the visual environment appears to be sensitive to knowledge of object affordances and local linguistic context: Listeners anticipate the location of expected referents and look to that region. In the paradigms reviewed above, different lexical items or different physical objects suggested different affordances that then influenced visual inspection patterns. How sensitive is inspection to even more subtle contextual information?

One interesting contrast in event descriptions may be illustrated by the following sentences: *He tipped the tea kettle.* versus *The tea kettle tipped.* In a context containing visual input of a tea kettle and either of these linguistic inputs, the surface form of the verb is constant as is the potential object affordance. However, in one description listeners receive information about an agent while in the other description no such information is provided. The presence versus absence of linguistic information about an agent may change people's inspection of the parts of objects that are especially associated with agents (e.g., *handles*, *steering wheels*, *triggers*). Borghi et al. (2004) suggested that, indeed, people are sensitive to the *parts* of objects that afford agentive action. How does this knowledge interact with particular frames of language to produce real-time visual search behavior?

Information search: A new use of eyetracking. In addition to studies of incremental sentence processing, eyetracking is a useful methodology to explore visual search in a broader sense. People explore the visual world for all sorts of reasons as they attempt to integrate information from multiple sources. In everyday experience, many things are often left unsaid. Sometimes this may occur because two interlocutors share common ground or a common visual context and do not need to reiterate shared information, while other times there is a genuine information gap at a particular time point during an interaction. In this latter case, visual search can sometimes lead to knowledge that fills this information gap.

One intriguing finding of this nature was recently reported by Crosby, Monin and Richardson (2006). In a novel eyetracking study of social referencing, participants watched a video of four people discussing affirmative action in university admissions. Three discussants were white and one was black. Crosby et al. found that when a white discussant strongly opposed affirmative action, listeners looked toward the black discussant. Crosby et al. suggested that people look toward the potentially-offended in potentially offensive situations to use their reaction as information about how to decide if discrimination had occurred and as cues for how to behave.

Even subtly different linguistic input conveys different information. In the case of agentive and non-agentive minimal sentence pairs, listeners receive information about the agent only from agentive sentences. Much of the previous eyetracking literature suggests that eye movements to visual scenes closely follow information in the linguistic input. On this account (and/or in situations in which this mechanism is most likely to be operative), *agentive* language might direct listeners' attention to regions in the visual world that are associated with agents. That is, after hearing agentive language such as "*He tipped the tea kettle*" listeners may look toward the interactive region of the object (i.e., the handle). On the other hand, a different mechanism of information search may operate when people attempt to elaborate their understanding of events. People may, in fact, search the visual world specifically for information that was *not* provided by language. In this case, *non-agentive* language may prompt people to fill an information gap by looking toward regions in which they expect to find agents. That is, after hearing "*The tea kettle tipped*", people may look toward the handle in order to learn about a potential agent. After briefly reviewing research examining some representational and processing consequences of agentive and non-agentive sentences, we introduce our study that aimed to examine whether, and how, these linguistic frames influence how people visually inspect objects with interactive regions.

Agentive vs. non-agentive language. What are the consequences of processing agentive and non-agentive language? Few psychological studies have addressed this question. One exception is Mauner and Koenig (2000), who compared the accessibility of agents in passive sentences such as *The baby's rattle was shaken repeatedly*, and active intransitive sentences such as *The baby's rattle had shaken repeatedly*. In English, passive sentences may be extended with agentive information (e.g., *by her mother*) while intransitive sentences may not. In a series of sentence processing experiments, Mauner and Koenig found that participants were quicker to detect contradictions between clauses that implied agents and passive sentences than these same agentive clauses and intransitive sentences. These results suggest that agents are less accessible after English intransitive sentences than after other types of sentences more strongly associated with the explicit linguistic encoding of agents.

Non-agentive language not only influences the salience of agents during language comprehension, but also influences learning and reasoning about agents and objects. For example, Fausey and Boroditsky (2007) found that people were sensitive to the distribution of agentive and non-agentive language that co-varied with observed events when learning about novel agents and objects. With increasing non-agentive language, people judged an agent to be less criminal and an object to be more capable of spontaneously transforming. In a separate series of studies that examined people's attributions of blame and financial penalties to causal agents, Fausey and Boroditsky (in preparation) found that people were more forgiving of agents of accidental events after reading descriptions that included non-agentive language (e.g., *The tablecloth ignited*) than after reading descriptions that included agentive language (e.g., *She ignited the tablecloth*). Agentive and non-agentive English sentences appear to influence a variety of reasoning behaviors. Does the reduced salience of agents after non-agentive language influence visual search of objects that afford agent action?

Present study. Previous research in the embodiment tradition suggests that the actions associated with particular objects partially constitute those object representations and that this knowledge is active during a variety of tasks. In the present study, we examined how local linguistic information interacts with global affordances knowledge to produce visual search behavior.

Linguistics research in this domain has generally focused on the incremental processing of active voice, transitive sentences. This type of language use is consistent with the affordance knowledge that people have about objects: Agents are explicitly mentioned (or are implied addressees in the case of imperatives) in the linguistic input and objects that afford human action appear in the visual input. What happens in the case of a mismatch - when people hear non-agentive language in combination with objects with agentive affordances – may inform broader questions about how people integrate information from a variety of sources.

We build on previous research that suggests that people are sensitive to the part of objects that afford action (e.g., Borghi et al., 2004) by examining people's eye movements toward "interactive regions" of objects (e.g., *handles, steering wheels* and *triggers*) following agentive versus non-agentive language. In each trial of our listen-and-look study, a sentence was presented auditorily, and then a static image of an object appeared on the screen. Eye movements were recorded for three seconds following the appearance of the image, and no overt responses were required of participants. Different visual inspection biases following different language may be revealed by this paradigm, extending both the eye movement and affordances literatures by providing evidence for how linguistic and more global knowledge interact during human information integration in service of event understanding.

Two potential outcomes may reveal sensitivity to object affordances in visual search. If language simply directs attention in this paradigm, participants should look toward the interactive region of objects more quickly following agentive language than following non-agentive language. That is, hearing "*He*" may lead people to look toward interactive regions of objects. It is also possible, however, that people will search for information that is *not* presented in the linguistic input. In particular, part of people's knowledge about the objects in our studies may be that agents typically cause the events in which the objects participate (e.g., *People* tip tea kettles). Upon hearing a non-agentive sentence such as *The tea kettle tipped,* participants may search for information about who tipped the tea kettle. That is, not hearing "*He*" may lead people to look toward interactive regions of objects.

How does the presence or absence of agent information in linguistic input influence subsequent visual search of objects with action affordances?

Present study: Inspected objects

When an object captures visual attention, where do viewers look first? Local contextual factors, as well as knowledge grounded in prior experience, indubitably determine the answer to this question. In this eyetracking study, English-speaking adults viewed objects with interactive regions, such as handles, knobs or buttons. Participants viewed each object after listening to a sentence with or without information about an agent. Participants' eye movements may inform questions about how global knowledge of object affordances and local linguistic information mutually determine visual inspection of objects.

Participants

Forty-five English-speaking students at the University of California, Santa Cruz, completed the study in partial fulfillment of a course requirement. All had normal or corrected-to-normal vision. We were unable to achieve a useful track on 16 participants due to equipment vagaries or vision correction (hard contacts or certain types of glasses).

Materials

Visual stimuli were 24 color photographs of objects against a white background, for instance, a *tea kettle*, *toilet* or *rifle*. Every object had an "interactive region": a clearly identifiable part that would be the site of any manual interaction, such as a handle, a button or a trigger. All images were 500 x 500 pixels, and interactive regions were determined with respect to each object. On average, the interactive region occupied 22 percent of the full image. Images subtended approximately 20° visual angle. See Figure 1 for an example visual stimulus.

Linguistic stimuli included 48 English sentences in the past tense. Two sentences were paired with each object: (1) one agentive sentence and (2) one non-agentive sentence. Agentive sentences were transitive sentences with the pronoun *he* (e.g., *He tipped the tea kettle*). Non-agentive sentences were intransitive sentences with no pronouns (e.g., *The tea kettle tipped*).

Figure 1: Visual stimulus used in the study.

All sentences were judged to be semantically acceptable (greater than a rating of 5.5 on a scale in which 1 was "definitely not English and unacceptable" and 7 was "definitely English and acceptable") by 17 English speakers who did not participate in the eyetracking study. Sentences were recorded by the same female native English speaker and presented aloud to participants. Table 1 lists all linguistic stimuli.[1]

Table 1: Linguistic stimuli.

Object	**Verb**	**Object**	**Verb**
bell	rang	balloon	lowered
jug	poured	sword	swung
pitcher	emptied	ketchup	squirted
rifle	fired	shirt	unzipped
gun	shot	tea kettle	tipped
faucet	shut off	hose	turned on
piano	played	car	started
maracas	shook	plane	flew
horn	blew	tractor	started up
truck	drove	blinds	shut
toilet	flushed	perfume	sprayed
racket	dropped	TV	turned off

Design

In the course of a passive listen-and-look task, participants were presented with 24 sentence-object pairs. Half the sentences were agentive and half were non-agentive, with agentivity assignment counterbalanced across participants. During each trial, participants first listened to a sentence while viewing a blank screen. Then, the object described in the sentence appeared. It remained onscreen for three seconds, during which time eye movements were recorded. Sentence-object pairs were presented in a random order, intermixed among other linguistic and visual stimuli.

[1] Table 1 includes verbs only. Participants heard full sentences that included each verb in an agentive frame or a non-agentive frame.
[2] This pattern was observed in each of the six consecutive 500ms windows comprising the viewing period.

Procedure

Participants were instructed to listen to sentences and to look at pictures on a screen. They were asked to pay attention to all stimuli, and told that they would not need to make any responses.

Participants completed the experiment in the Eye Think lab's (D.C.R.) speech and gaze tracking system. Each participant sat in a reclining chair, looking up at an arm-mounted 19" LCD screen approximately 24" away. A Bobax3000 remote eye tracker, consisting of a camera focused on the participant's eye and a set of LED illuminators, was mounted at the base of the display. Each participant wore a headset, through which s/he listened to stimuli.

Intel iMacs were used to present stimuli and to record data. The eye trackers passed image data to the iMacs, which calculated gaze position for each participant approximately 30 times a second and recorded regions of interest that were being fixated. Data were also streamed to an experimenter's computer, which saved an audio-video record of what participants saw, heard and said during the experiment, superimposed with their gaze position.

Results

Participants viewed object images for the final three seconds of each trial (3000ms to 6000ms), and eye movements were recorded for this time period.

Participants spent similar total amounts of time looking at the interactive region of objects following non-agentive (M = 867 ms) and agentive language (M = 894 ms), $F(1, 28) = .20$, $p = .66$.[2] However, participants were approximately 100 milliseconds *faster* to fixate the interactive region of objects after hearing non-agentive language (M = 3435 ms, where the onset of the picture is at 3000ms) than after hearing agentive language (M = 3543 ms). This pattern was reliable across participants, $F(1, 28) = 4.72$, $p = .038$, as well as across items, $F(1, 23) = 6.25$, $p = .02$ (see Figure 2).

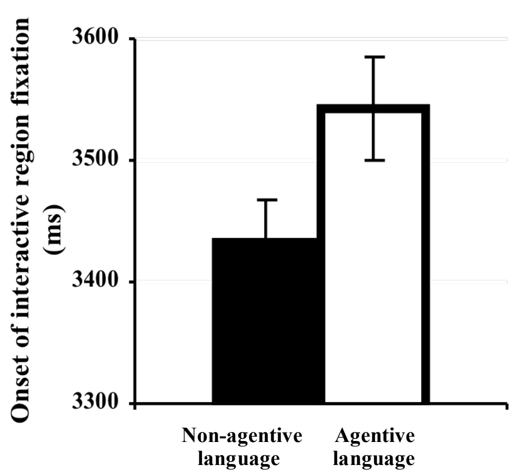

Figure 2: Onset of interactive region fixation

Discussion

Participants viewed objects with affordances for human action after listening to simple sentences. People's visual inspection of these objects was influenced by information in the sentence: Participants looked toward the interactive region of the object more quickly after sentences that did *not* mention an agent (e.g., *The tea kettle tipped*) than after sentences that did mention an agent (e.g., *He tipped the tea kettle*).

At first blush, this finding may seem surprising. Most previous studies using incremental processing paradigms have shown that listeners look at what is talked about. One might expect, then, that hearing "*he*" in agentive sentences would direct listeners' eye movements to the regions of objects associated with agents. However, in our anticipatory paradigm, the lack of agentive information in the linguistic input directed listeners to quickly fixate the interactive regions of objects. One intriguing explanation for our results may be that people expect agents to do things like *tip tea kettles, flush toilets* and *fire guns*. When no information about agents is provided by language, listeners may attempt to "fill" this information gap by quickly fixating to the interactive region of objects.

Some evidence in support of this explanation comes from a subsequent eyetracking study in which we examined people's visual inspection of objects after hearing two different kinds of agentive sentences. This study was identical to the present study, with two exceptions: (1) At the beginning of the experiment, participants viewed a photograph and listened to short biographical statements about three men (*Bill*, *Dave* and *Tom*). For example, participants were introduced to Bill, hearing "*This is Bill. He's 23 and likes history*", and (2) During the listen-and-look procedure, participants heard 12 sentences starting with an agent name (four sentences per name), such as "*Bill tipped the tea kettle*", and 12 sentences starting with an agent pronoun, such as "*He tipped the tea kettle*". Thus, participants heard agentive sentences in both conditions but received less information about the agent from the pronoun sentences than from the name sentences. With this more subtle manipulation of the amount of linguistic information about agents, the type of sentence did not influence how *quickly* people looked toward the interactive regions of objects. It did, however, influence whether or not people looked at these regions at all: People looked at the interactive regions of objects more after hearing the pronoun than after hearing the specific agent name, $t(24) = 2.08$, $p = .049$. Though both "*Bill tipped the tea kettle*" and "*He tipped the tea kettle*" are plausible sentences that explicitly mention an agent, people looked to interactive regions of objects more often after receiving the less informative "*he*", as if searching for more information about the agent.

In addition to information search, simulation mechanisms may play a role in how people integrate visual and linguistic input. Non-agentive sentences may create stronger pressure to simulate events than do agentive sentences. In the absence of both a linguistic agent and a visual agent, people may need to simulate the event as if they were the agent in order to understand the event. This simulation might lead to increased speed or amounts of looking to the interactive regions of objects as people imagine interacting with the object. Non-agentive sentences in our paradigm may have been most efficiently processed using a simulation mechanism, but note that in our follow-up study, information *per se* seemed to guide people's inspection of the interactive regions of objects.

Agentive and non-agentive sentences differ in several ways, all of which may contribute to how people visually inspect objects that afford action. Not only do agentive and non-agentive sentences contain different amounts of information about agents, but they also seem to convey different information about motion and end-states. In a paper-and-pencil study, 300 UC Merced students were presented with an agentive or a non-agentive sentence from our eyetracking studies and were asked to draw a picture of what came to mind. As expected, participants drew agents more often when depicting agentive sentences than when depicting non-agentive sentences, $\chi^2(1) = 133.02$, $p < .001$. Additionally, a naïve coder rated the degree of motion (e.g., motion lines) in each drawing, using a four-point scale from "none" to "high". Analyses of these ratings revealed that participants drew more motion in non-agentive depictions ($M = 2.13$) than in agentive depictions ($M = 1.76$), $t(298) = 2.53$, $p = .012$. Finally, though most participants depicted events at the midpoint of a "beginning-middle-end" timeline, participants were more likely to draw beginning states when depicting agentive sentences and more likely to draw end-states when depicting non-agentive sentences, $\chi^2(2) = 21.50$, $p < .001$.

Related research on mental imagery and language suggests that people naturally direct their attention to imagined changes in location without any visual stimulus (e.g., Richardson & Spivey, 2000; Spivey & Geng, 2001) and that they imagine movement and scan along paths when processing sentences that include fictive motion, such as *The road goes through the valley* (Matlock, 2004; Richardson & Matlock, 2007). Our drawing results suggest that people may imagine different scenes after agentive and non-agentive sentences, and this different imagery may impact subsequent visual search.

A number of future directions will help to unravel the many mechanisms that contribute to the integration of global affordance knowledge and local linguistic information in the visual inspection of objects. Because the current study was designed to specifically examine interactive regions of objects, not all visual stimuli had clearly identifiable "end-state" regions. To better understand how people integrate agentive and non-agentive linguistic frames with visual scenes, we are currently extending our paradigm by presenting people with images that depict agents (e.g., a person), objects (e.g., a tea kettle) *and* end-states (e.g., a puddle of water). Explicit visual depictions of all aspects of the event may help to more precisely understand how agentive and non-agentive language are integrated with knowledge of object affordances during visual processing. Future research may also consider additional linguistic manipulations, such as a no-language baseline, as well as contrasting active vs. passive constructions.

One particularly interesting future direction with respect to linguistic manipulations would be to examine the integration of language, visual processing and knowledge about object affordances in speakers of languages other than English. Because there is cross-linguistic variation in the distribution of agentive and non-agentive expressions in a language (see Fausey & Boroditsky, 2006), people in different language communities may vary in their need to fill agent information gaps. For example, if a community of speakers commonly talk about *tea kettles tipping*, *toilets flushing* and *guns firing*, without mentioning agents, perhaps non-agentive language would not so strongly bias looking toward interactive regions of objects. Using this paradigm to examine visual inspection patterns of people in different linguistic communities may help us to better understand how language is integrated with other knowledge to constrain processing of objects and events.

Conclusion

When visually inspecting everyday objects, participants were faster to fixate the interactive region of these objects after hearing non-agentive language than after hearing agentive language. This preliminary research is suggestive of an influence of linguistic framing on visual information search. Future research will continue to elaborate mechanisms by which people integrate their rich knowledge of agents, everyday objects and language as they visually explore their world.

Acknowledgments

We thank Michael Spivey and Alexia Toskos for helpful discussions of this work. We thank Jason Hreha for help in preparing the stimuli, Arezou Ghane for help with data collection, and Vijay Vanchinathan for coding drawings.

References

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

Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition*, *32*, 863-873.

Chambers, C. G., Magnuson, J.S., & Tanenhaus, M. K. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory & Cognition, 30*, 687-696.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage, 12*, 478–484.

Creem, S.H., & Proffitt, D.R. (2001). Grasping objects by their handles: A necessary interaction between cognition and action. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 218-228.

Crosby, J.R., Monin, B., & Richardson, D.C. (2006). Looked at, but not listened to: Focusing on the reaction of minority group members when deciding if discrimination has occurred. Paper presented at the *7th Annual Meeting of the Society for Personality and Social Psychology*.

Fausey, C.M., & Boroditsky, L. (in preparation). Speaking of accidental agents.

Fausey, C.M., & Boroditsky, L. (2007). Language changes causal attributions about agents and objects. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Fausey, C.M., & Boroditsky, L. (2006). Linguistic contributions to reasoning about causal agents. *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum

Glenberg, A.M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*, 1-55.

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

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133-156.

Mauner, G., & Koenig, J-P. (2000). Linguistic vs. conceptual sources of implicit agents in sentence comprehension. *Journal of Memory and Language, 43*, 110-134.

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

Pecher, D., & Zwaan, R.A. (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge: Cambridge University Press.

Richardson, D.C., & Spivey, M.J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition, 76*, 269-295.

Richardson, D.C., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition, 102(1)*, 129-138.

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

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 830–846.