



# Chapter 11

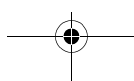
## The On-Line Processing of Ambiguous and Unambiguous Words in Context: Evidence from Head-Mounted Eyetracking

FALK HUETTIG AND GERRY T. M. ALTMANN

### 11.1 BACKGROUND: THE INTERACTION OF LANGUAGE, ATTENTION, AND VISION IN REAL-WORLD BEHAVIOR

A main goal of the study of language processing must be to understand how language processing affects aspects of human behavior. Traditionally, language processing has almost exclusively investigated language in isolation from the environment in which it is usually used. Although one of the defining features of language is that it can be used to refer to objects in their absence, language *is* often used to refer to objects in the immediate environment. From an evolutionary perspective, the sensory environment is likely to have had (and has) an important impact on the evolution of language and, more generally, cognition. Similarly, from a developmental perspective, language acquisition appears to be reliant in large part on a concurrent world with which the infant/child can interact. Notwithstanding this reliance on an external world (and even in its absence language does *refer* to objects and events in that external world), there has been little research that directly explores the interface of language and visual perception and its impact on our interactions with the visual world (but see Henderson & Ferreira, 2003, for reviews of such work).

Many theoretical accounts of language processing have assumed modular semantic or conceptual systems for the representation of knowledge that are theoretically (and possibly physically) distinct from the episodic modality-specific systems that support perception and action. In contrast, recent theories of *embodied* cognition (e.g., Barsalou, 1999; Glenberg, 1997) propose that conceptual representations are grounded in the neural substrate that supports interaction with, and perception of, the external world. Common to all embodied theories of cognition is a rejection of the claim that the human representational system is independent of the biological systems that embody it. This is in contrast with the view that cognition and knowl-



## 188 THE ON-LINE STUDY OF SENTENCE COMPREHENSION

edge can be considered independently of perception and action. Barsalou proposes that the meaning of a word is, in effect, a context-sensitive representation which, when activated, causes the reenactment of previous experiential states. These “simulations” develop through repeated experience of objects, actions, mental states, and so on (see Barsalou, Simmons, Barbey, & Wilson, 2003, for review). The experience may be perceptual or sensory, or motoric, and thus the cognitive representations that ensue are embodied in the neural substrate that supports such experiences. When a word is recognized, the cognitive system activates the simulation of the associated concept, and this may include the reenactment of experiential states that include sensory and motoric components.

*Situated* accounts of cognition fit well with embodied accounts. Situated cognition refers to theories that propose that the perceptual system “offloads” information by leaving it in the environment (or the “external world”) rather than just passively passing information on to the cognitive system for propositional (*amodal*) representations to be created. According to this view, perceptual information in the environment is accessed when needed, with the visual world functioning as a kind of external memory (e.g., O’Regan, 1992). Objects in this situated memory are represented in a spatial data structure which contains “pointers” to the real-world location of the object. Thus, the system need not store internally detailed information about the object, but can instead locate that information, when it has to, by directing attention back to that object in the external world. Note, however, that we assume situated memories to be content-addressable and to contain some information about objects to permit content-addressability, i.e., just enough information to allow the system to acquire the detailed information that was not stored.

In sum, a word’s meaning must be represented, according to the embodied cognition approach, across several sensory modalities. In effect, its meaning is composed of different sensory impressions abstracted across multiple experiences of that object in multiple modalities. If one receives a combination of sense impressions with any degree of frequency, simple principles of learning (Hebbian learning, for example) predict that, subsequently, any single member of the combination could trigger the others (see also Calvin & Bickerton, 2000). Embodied and situated theories of cognition propose, therefore, that cognition is perceptually grounded, and that conceptual processing involves perceptual simulation rather than the manipulation of the abstract and arbitrary symbols that are assumed in traditional cognitive science (see Barsalou et al., 2003, for a brief overview of this position).

### 11.2 THE IMMEDIATE BEHAVIORAL CONSEQUENCES OF THE DISPLAY OF PERCEPTUAL COMPETITORS

If conceptual representations are abstractions across modality-specific experiences, and if language comprehension involves perceptual simulation (the reenactment of embodied experiential states), then variables such as perceptual similarity should affect language-mediated eye movements in our interactions in the visual world. For example, hearing a word such as *cabbage* should give rise to an experiential state that reflects the perception of properties such as roundness and greenness, as well

EVIDENCE FROM HEAD-MOUNTED EYETRACKING **189**

as other properties stemming from a cabbage's typical smell and taste. The reenactment of these experiential states (i.e., the activation of the *perceptual features*) will direct visual attention to objects in the environment which share these same features. (See the following text for an account of the mechanism by which visual attention is directed to these locations.) For the moment, we simply assume that a mechanism *is* in place to direct attention toward a cabbage when the word *cabbage* is heard. The system must presumably match the *simulated properties* against whatever perceptual properties are afforded by the objects in the visual scene, and thus objects that share some (but not necessarily all) of those perceptual properties may attract visual attention more than others that do not share those properties. The experiments that follow explore this prediction. Specifically, we ask whether hearing a word such as *cabbage* causes more eye movements toward objects that share physical shape (roundness) or color than toward objects that do not. Given that the eyes *can* be directed towards cabbages when we hear *cabbage*, we must be matching the perceptual features recovered through hearing *cabbage* against the perceptual features afforded by the objects in the visual environment. What is less clear is whether, when we have identified those objects, we direct our attention on hearing *cabbage* to any of them that, although not cabbages, are nonetheless round or green.

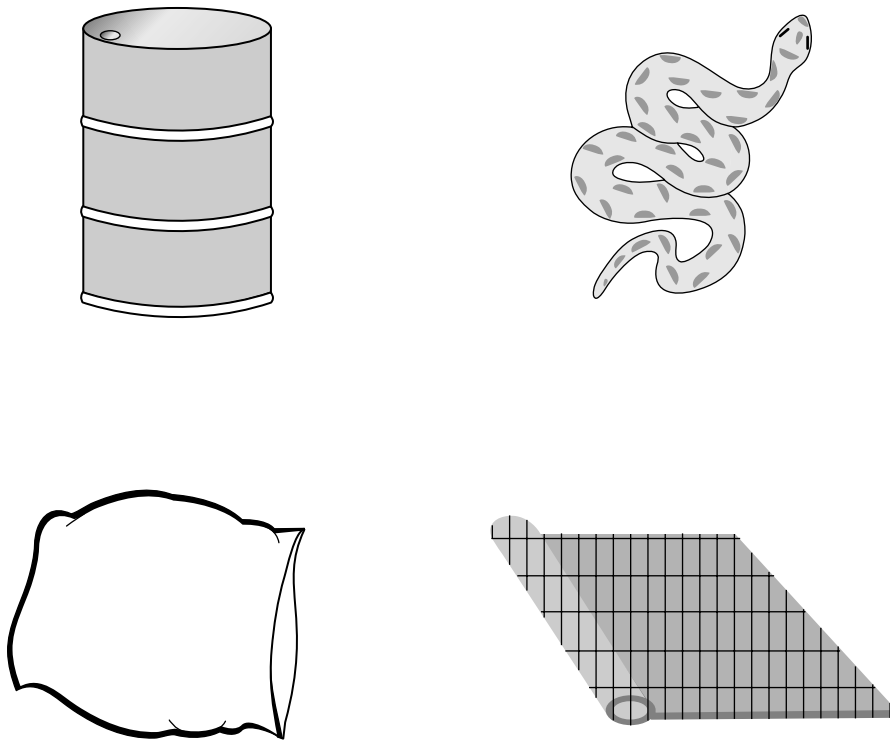
In the studies that follow, participants heard sentences such as *In the beginning, the man watched closely, but then he looked at the snake and realized that it was harmless*. Simultaneously, they viewed a visual scene depicting four objects. The target word, in this example, was *snake*, and the delay between the onset of the sentence (and the concurrent onset of the visual scene) and the onset of this target word, was approximately 5 s. We asked our participants to listen to the sentences carefully. They were told that they could look wherever they want but we asked them not to take their eyes off the screen (i.e., the listeners received no instructions other than to listen to the sentences carefully). We recorded participants' eye movements as they listened to the sentences.

### 11.2.1 Experiment 1: Visual Form Similarity

In the first study, we created two sets of visual stimuli. The stimuli in one set (the "target" set) consisted of a picture depicting an unambiguous target word (e.g., a snake) and three pictures depicting objects from different semantic/conceptual categories than the target word. These distractor items were chosen so that their names were frequency-matched with the target according to the CELEX (Baayen, Piepenbrock, & van Rijn, 1993) database. The stimuli in the second set (the "visual competitor" set) consisted of the same four pictures in the identical positions except that the target picture (the snake) was replaced with a picture depicting an object with a similar visual form as the target word (e.g., a cable). See Figures 11.1 and 11.2 for examples.

We constructed 21 sentential stimuli in two versions: a neutral context condition and a biasing context condition. For the neutral context condition, the sentence did not provide any contextual bias up until the target word that would favor any of the pictures depicted in the visual scene: *In the beginning, the man watched closely, but then he looked at the snake and realized that it was harmless*. For the biasing context

## 190 THE ON-LINE STUDY OF SENTENCE COMPREHENSION

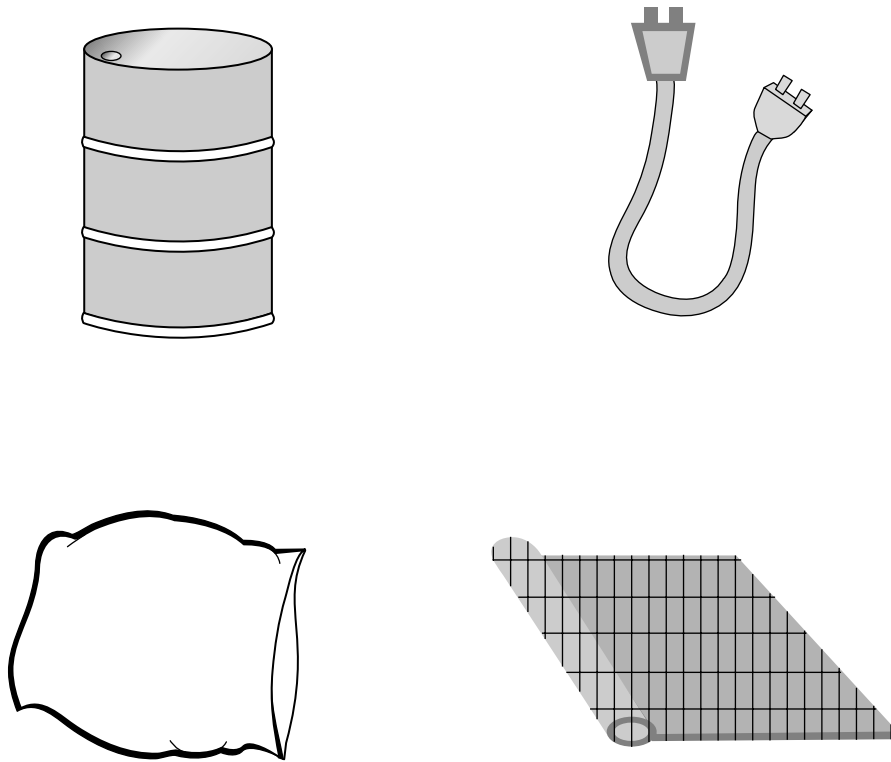


**Figure 11.1** An example of a visual stimulus for the target set.1

condition, the sentence biased toward the target object (the snake): *In the beginning, the zookeeper worried greatly, but then he looked at the snake and realised that it was harmless.* These sentences were identical to the neutral condition except for a single phrase that biased the upcoming target word (*zookeeper worried greatly*) which replaced the neutral phrase in the neutral condition (*man watched closely*). In the experiment itself, there were three conditions: the neutral sentences with the target stimuli, the biasing sentences with the target stimuli, and the biasing sentences with the visual competitor stimuli. Our rationale for presenting the visual competitor in a biasing context was simply that we wanted to make it relatively unlikely that participants would anticipate, prior to the target word, that the visual competitor would be the object of attention (even though it was not going to be referred to directly). The neutral sentences were included in order to establish a baseline against which the efficacy of the biasing context could be determined; the idea here was that in the neutral context there would be no advantage (in terms of attracting looks) of the target object until the corresponding target word was heard, but in the biasing context (if the attempt to induce a bias were successful), an advantage for the target object should be observed prior to the target word. We return to discussion of these contexts below.

In these first studies we report the percentage of trials in which a saccade was launched towards the target object during the acoustic lifetime of the target word,

## EVIDENCE FROM HEAD-MOUNTED EYETRACKING 191



**Figure 11.2** An example of a visual stimulus for the corresponding visual competitor set.

AU: "Altmann & Kamide, in press" not included in ref list.

and, where appropriate, the percentage of trials on which the different objects were being fixated at the acoustic onset of that word.<sup>1</sup> The latter measure gives an indication of the bias to look toward one object or another before any of the acoustic input pertaining to the target word has been encountered.

In the neutral context condition, there was no significant bias at target word onset to look at the snake any more than at the distractors. During the target word (which averaged 447 ms in duration), participants initiated saccadic eye movements toward the target object (in fact, toward the quadrant within which the target object was depicted) on 35% of trials. Saccades toward the distractors averaged 10% of trials (that is, the three distractors each attracted saccades on 10% of trials). This difference was highly significant in statistical analyses by subjects and by items.<sup>2</sup> In the biasing context condition, there was a significant bias, at target word onset, to fixate *the snake* more than the distractors (target, 32%; distractors, 21%). During the target word, saccades were initiated toward the target object on 25% of trials, and to the distractors on 10% of trials. Again, this difference was highly significant. Although it may seem counter-intuitive that we see *fewer* looks toward the target object when the context biased towards this object, this reflects, simply, the increased likelihood of already fixating *the snake* in the biasing context. The fact that there was a significantly increased bias at target word onset to fixate *the snake* in the

## 192 THE ON-LINE STUDY OF SENTENCE COMPREHENSION

biasing context compared to the neutral context (biasing context, 32%; neutral context, 21%) suggests that our contextual manipulation had been successful (although we are less concerned with the magnitude of the bias than with the fact that there is a bias at all). Of critical interest were the data from the visual competitor condition in which *the snake* had been replaced by a cable. At the onset of *snake*, the cable and each of the distractors were being fixated with equal probability. However, postonset, during the lifetime of *snake*, more saccades were initiated toward the cable (35% of trials) than toward any of the distractors (each distractor attracted saccades on 18% of trials). Thus, as *snake* unfolded, participants executed more saccades toward the visual competitor (*the cable*) than to the visually unrelated distractors.

In summary, the patterns of eye movements to the visual competitor closely resembled the eye movements to the target object depicting the target word and did so from the earliest moments in which information from the unfolding target word (*snake*) became available (cf. Dahan & Tanenhaus, 2002; Dahan & Tanenhaus, 2002). AU: Update? **submitted** although nothing in the linguistic context biased toward the visual competitor. Thus, visual attention during real-time speech processing seems to be directed toward objects in the visual world that match the physical shape of whatever is being referred to by the language. Importantly, the acoustic onset of the target word in these studies occurred on average 5 s after the onset of the visual scene. That is, any eye movements initiated toward *the snake* in response to *snake* most likely were initiated toward that location because participants already knew there was *snake* at that location. Similarly, any eye movements initiated toward the cable in response to *snake* most likely were initiated toward that location *despite* participants knowing that there was a cable at that location.

### 11.2.2 Experiment 2: Color

Our next study explored whether we could extend this finding to perceptual properties other than shape, such as color. Participants heard sentences such as *The boy turned around carefully, and then he saw the frog and looked happy*, which were designed to be linguistically neutral with respect to any of the visually presented pictures. The concurrent visual stimuli consisted of sets of four line drawings taken from Snodgrass and Vanderwart (1980). In the target condition, the four line drawings included the target object (e.g., a frog) and three distractors that were unrelated conceptually to the target (e.g., a mitten, a pipe, and a suitcase). The target was colored according to its prototypical color (green, in this case), and the distractors were colored in ways that were deemed normal (e.g., a blue mitten, a brown pipe, and a red suitcase). In the color competitor condition, the target object was replaced by a color competitor: for the frog, the color competitor was a lettuce. The procedure in this study was the same as in the previous one. The data were similar: significantly more saccades were initiated during the acoustic lifetime of the target word towards the competitor than towards the other distractors. Thus, in the target condition, the frog engendered saccades on 52% of trials, and each distractor on 12% of trials. In the visual competitor condition, the lettuce engendered saccades on 41% of trials, and the distractors on 26% of trials. There were no advantages of either the target

or the competitor prior to the onset of the target word. Thus, participants initiated significantly more saccades during the acoustic lifetime of the target word *frog* toward the picture of the color competitor (*lettuce*) than toward the distractors (*mitten*, *pipe*, and *suitcase*).

These data, together with those from Experiment 1, demonstrate that language-mediated visual attention is directed immediately toward objects that match on the perceptual properties of form and color. We suggest that information activated during the acoustic lifetime of the target word (including information about visual form and color) causes the activation of stored representations of the visual objects that share such information. Why this engenders immediate looks to those objects will be left until later discussion. First, we consider whether similar effects can occur if the overlap between the target word and the object in the visual field is based not on perceptual properties of real-world objects but on their semantic properties.

### 11.3 THE IMMEDIATE BEHAVIORAL CONSEQUENCES OF THE DISPLAY OF CONCEPTUAL COMPETITORS

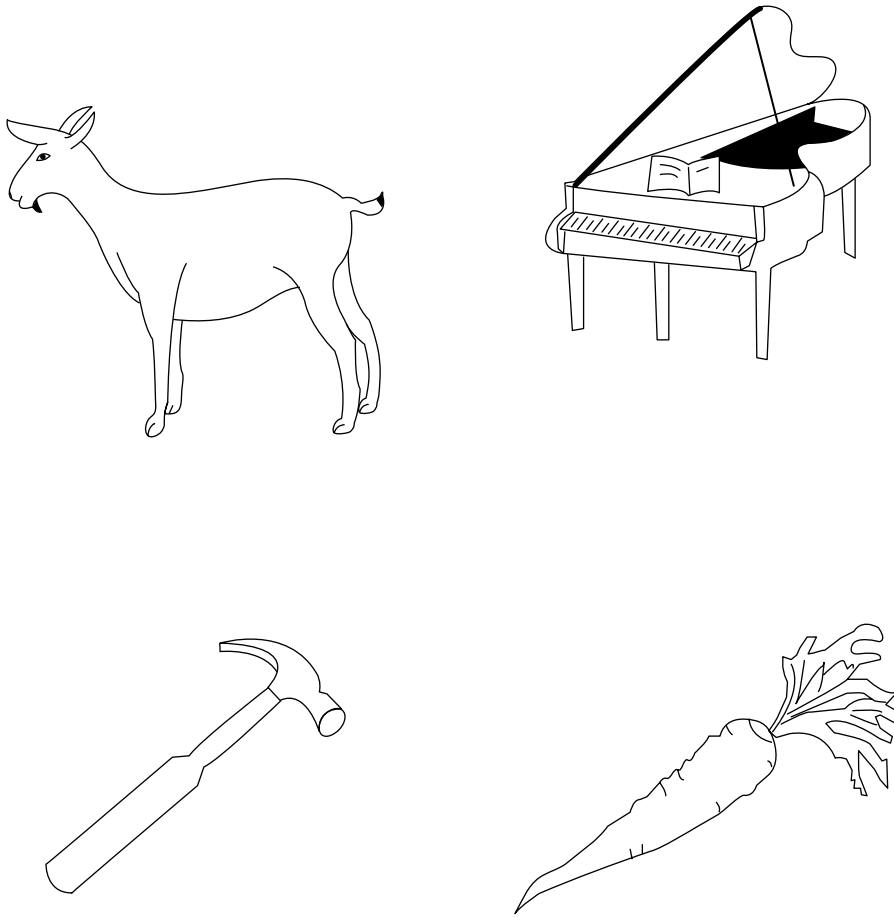
#### 11.3.1 Experiment 3: Conceptual Category

Are the effects we have observed thus far confined to perceptual features or might they generalize to *conceptual* features also. Work by Eiling Yee and Julie Sedivy suggests that it may: they observed increased looks toward a key when the word *lock* was heard (Yee & Sedivy, 2001). In their study, the lock was co-present in the visual scene, and it is possible that looks to the key were initiated once the lock had been attended to. However, like them, we suspect that their finding was the visual equivalent of semantic priming (in this example, mediated by a *functional* relationship between the lock and the key). Our next study was designed to investigate such effects further.

The basic design consisted of three alternative visual scenes each accompanied by the same target sentence: *Eventually, the man agreed hesitantly, but then he looked at the piano and appreciated that it was beautiful*. The target word was *piano*, and the three alternative scenes differed in respect of the objects they portrayed (see Figures 11.3 to 11.5). In each case, the scene was composed of four line drawings. In the first condition, one drawing depicted the referent of the target word (e.g., a piano), and the remaining three drawings were distractors that were unrelated in visual form, prototypical color, or conceptual category. In the second condition, the target from the first condition was replaced by a conceptual competitor (a trumpet), but the three distractors remained the same. The third condition portrayed both the target and the conceptual competitor, as well as two distractors (cf. the Yee and Sedivy study, but see the following text).

The pictures were black and white line drawings from the normed Snodgrass and Vanderwart (1980) set and were matched for picture-naming agreement, familiarity, frequency (of the corresponding name), and other variables. All the pictures for each item were from different conceptual categories. Similarly to Snodgrass and Vanderwart we used the Battig and Montague (1969) category norms as a guide. We used only pictures from the Snodgrass and Vanderwart set that were also mem-

## 194 THE ON-LINE STUDY OF SENTENCE COMPREHENSION



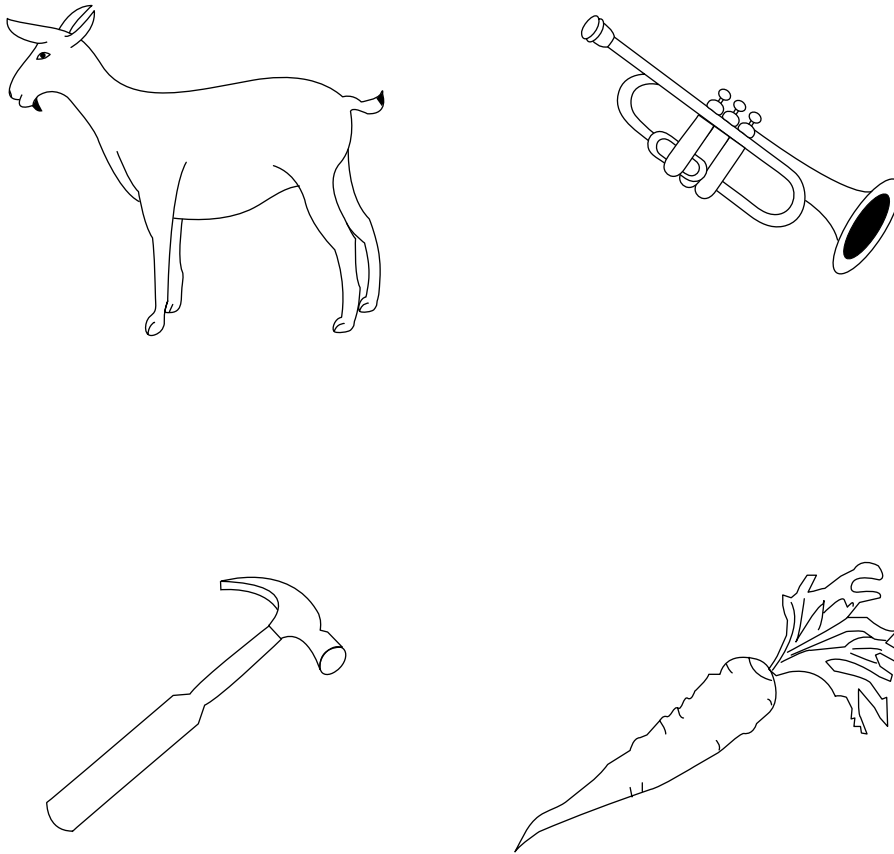
**Figure 11.3** An example of a visual stimulus in the first condition (target: *piano* depicted).

bers of the Battig and Montague norms. We selected pictures from the following conceptual categories: four-footed animal, furniture, human body part, kitchen utensil, musical instrument, clothing, type of vehicle, part of building, weapon, fruit, carpenter's tool, bird, toy, insect, and vegetable. None of the conceptual competitors were associates of the target words (unlike the lock and key example used by Yee & Sedivy, 2001). We consulted the University of South Florida word association norms (Nelson, McEvoy, & Schreiber, 1998) that typically used hundreds of participants in the collection of the norms for each word. We used a stringent exclusion criterion: if even a single participant produced the competitor after the target, or vice versa, we rejected that item. For example, if only one participant of the several hundreds in that norming study had produced *trumpet* after *piano* or *piano* after *trumpet*, then we would have excluded this item.

We analyzed our data in the same way as in the previous experiments. In the first condition, with just the target (the piano) and three distractors, there were more



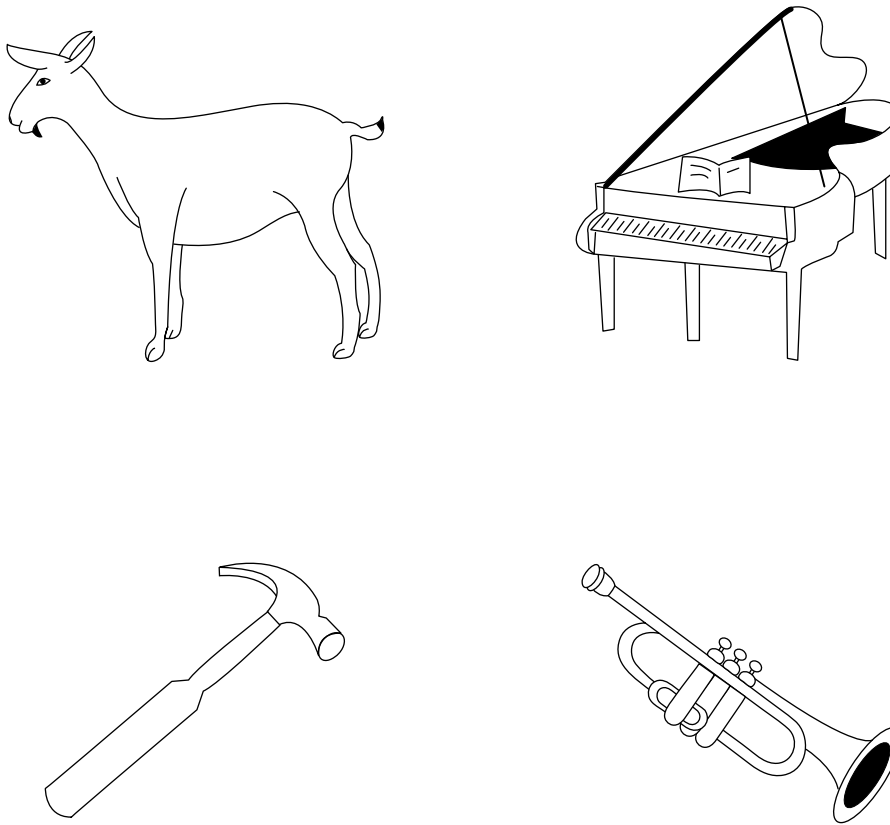
## EVIDENCE FROM HEAD-MOUNTED EYETRACKING 195



**Figure 11.4** The visual stimulus in the second condition (target picture replaced by a picture of the category competitor *trumpet*).

trials with looks initiated toward the piano (39%) than with looks initiated toward the distractors (14% each). In the second condition, which portrayed instead the conceptual competitor (the trumpet), there was a large difference, during the target word, between looks toward the trumpet (36%) and looks to each of the distractors (19%). In both conditions there were no biases by the onset of the target word that favored looks toward any of the drawings (whether target, competitor, or distractor)—all were being fixated at this point with equal probability. Thus, hearing *piano* engenders immediate looks toward the conceptually related trumpet. The data from the third condition were a little more complex because the visual scene portrayed both the target and the competitor and, unsurprisingly, there were more looks towards the target than toward the competitor (41% vs. 17%) but more looks to both of these than to the two distractors (9% each). The likelihood of the distractors being fixated at the onset of the target word was marginally greater than that for the target or conceptual competitor being fixated (27% for the distractors, and 22% for each of the target and conceptual competitor); thus, although it was marginally less likely

## 196 THE ON-LINE STUDY OF SENTENCE COMPREHENSION



**Figure 11.5** The visual stimulus in the third condition (target *and* competitor depicted).

at the onset of *piano* that the target and competitor were already being fixated, there were more looks initiated post-onset toward the target and competitor than toward the distractors. The greater advantage of the target relative to the conceptual competitor is accounted for by the fact that the conceptual competitor (*trumpet*) only matched on basic conceptual dimensions, while the target (*piano*) matched on both conceptual and perceptual properties.

Taken together, these data demonstrate that visual attention is directed immediately, as a word unfolds, toward objects that match the target specification of that word on the grounds of conceptual category.<sup>3</sup> The fact that the pictures corresponding to these objects had been on-screen for approximately 5 s suggests that our effects were not due to any confusion regarding the identities of the four portrayed objects. We note in this regard that Dell'Acqua and Grainger (1999) observed unconscious activation of semantic information from picture stimuli after exposures of just 17 ms. Interestingly, visual attention appears to be directed toward conceptually related objects even if they mismatch on other grounds (perceptual shape and color, and conceptual detail). Finally, attention is directed towards such mismatching objects even when an object exists in the visual field which meets the full conceptual and

perceptual specifications associated with the critical target word. Before considering the mechanism which causes such effects, we shall turn to one final question that arises following the conceptual competitor data. They concern the locus of the color effect observed earlier.

### 11.3.2 Experiment 4: Putting Color in Black and White

Why did participants look at the lettuce when they heard *frog*? Is it because they accessed the color information associated with frogs, and their eyes were then attracted to anything that they *saw* was green? Or were their eyes attracted to anything that they *knew* was green? In other words, is the color competitor effect a *semantic* effect, or a *perceptual* one? To explore this, we repeated the logic of the color competitor experiment, but in black and white. Participants were shown a four-picture scene with, for example, a frog, a mitten, a pipe, and a suitcase. We again used line drawings from the Snodgrass and Vanderwart (1980) set. Participants heard either *The boy turned around carefully, and then he saw the frog and looked happy*, or *The boy turned around carefully, and then he saw the spinach and looked happy*. In the first case, the visual scene affords a referent for the target word *frog*, but in the second case, there is no referent for *spinach*, although there is a conceptual color match between spinach and the frog portrayed in the scene. In the *frog* condition, participants initiated a saccadic movement toward the frog, during *frog* on 52% of trials. Each distractor attracted saccades on 12% of trials. In the *spinach* condition, on only 30% of trials did participants look toward the frog during *spinach*. And in this case, the distractors each attracted saccades on 27% of trials. Thus, there was no significant difference between looks to the frog and looks to the distractors in this condition (the numbers add up to more than 100 because there were some trials on which there was more than one fixation within the lifetime of the target word). The difference in looks as a function of the target word (*frog* vs. *spinach*) cannot be explained by any difference in the duration of the two words, as these averaged 480 and 490 ms, respectively.

These data suggest strongly that the color competitor effect is driven by *perceptual*, not *conceptual*, factors (given that prototypical color could be deemed to be a conceptual or semantic feature). In the following section, we offer a tentative explanation of our competitor effects before proceeding to the application of these effects to the study of lexical ambiguity.

## 11.4 LANGUAGE-MEDIATED EYE MOVEMENTS, “EMBODIED LOCATION,” AND COMPETITOR EFFECTS

Why do the eyes move to form, color, or conceptual competitors? And why do they move so quickly? In respect of conceptual competitors, proponents of situated vision would argue, presumably, that the conceptual information that is accessed (or activated) on hearing the target word causes the eyes to move towards the depicted conceptual competitor in order to retrieve the situated memory of what is at that location. In other words, the system has a record that something with particular conceptual features is located at that location, and when the target word is heard,

PE: Author had these words in italic, enclosed in quotes. We have removed the quotes.

**198** THE ON-LINE STUDY OF SENTENCE COMPREHENSION

the system needs to retrieve information about that object in order to establish its fit with the target specification as provided by the target word. But if this is the case, what is the *mechanism* by which attention is directed toward that location? The situation in respect of situated vision and the perceptual competitor effects (of form and color) is a little more complex; the word *frog* makes available prototypical color information, and situated memories cause the eyes to move toward something sharing that color in order to recover further details. However, the conceptual competitor effects suggest that enough is known about whatever is in the location to which the eyes will be directed to know that it is of a certain conceptual type (if *piano* causes us to attend to the trumpet, this must be because we know that *piano* refers to a musical instrument, and that a musical instrument is located at the position in the visual field that is in fact occupied by a trumpet). So why, when we hear *frog*, should we move to the lettuce when, evidently, we know that *what we want* is animal-like, and that *what we will get* is vegetable-like?

The answer to these puzzles is to consider how information about location might be encoded within the cognitive system. Within the embodied approach to cognition, the representations of objects in the world are encoded in the same substrate whose activation has supported the experience of those objects (cf. Glenberg, 1997). Presumably, the representation of the *location* of those objects would be encoded in the same substrate whose activation supported the experience of those objects *as mediated by* physical movement with or toward those objects (whether in terms of orientation toward the objects or shifts in visual attention toward those objects). Thus, if one attends to an object, the neural substrate implicated in the corresponding shift in attention towards that object encodes that object's location. The relevance of this embodiment is in terms of what happens when the representation of an object is subsequently reactivated. Activating the representation of an object necessarily activates the embodiment of the experience of that object, and because this experience necessarily includes an attentional component, the embodiment of that attentional component is also activated. We conjecture that reenacting this component will, in the absence of any competing attentional demands, cause the eyes to move automatically toward that object's location. The claim, then, is that the activation of the mental representation of a specific object in the visual scene will automatically cause a shift in attention towards the location at which that object was located (at the time that it was experienced). This in itself does not explain our competitor data. However, little more is required to explain those effects than the same processes implicated in priming. In respect of semantic priming (e.g., McRae & Boisvert, 1998; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995), it is commonly supposed that the activation of a particular semantic feature causes the activation of other concepts that share that feature (where "semantic feature" is shorthand, within the embodied view of cognition, for a particular component of the neural substrate that encodes the relevant abstractions across the multiple experiences that led to the formation of the concept). In respect of our conceptual competitor data, we can suppose that hearing *piano* activates concepts that share experiential features, and in this case, experiential features that correspond to the (emergent) superordinate level category information associated with pianos (i.e., that they are musical instruments). The conceptual representation corresponding to the trumpet shares such information and

## EVIDENCE FROM HEAD-MOUNTED EYETRACKING 199

has associated with it the episodic information corresponding to its experience previously in the visual scene. The ensuing change in activation state of this representation causes the reenactment of the experiential state associated with the trumpet, and hence (according to the previous arguments) the shift in attention toward it.<sup>4</sup>

In respect of the color competitor effects, a similar chain of events is required with one subtle, but as we shall see, perplexing, difference: Hearing *spinach* activates information about the prototypical color of spinach, but it is not information about the prototypical color of the visually presented frog that attracts looks; instead, it is information about the actual colors in the visual scene that attracts looks. In other words, activation does not spread to all things that are prototypically green. If it did, and by extension of the previous arguments, we should have seen looks toward the black and white frog in response to *spinach*. But we did not. Instead, activation of the color information associated with spinach results in orientation towards the location associated with the *perception* of color, rather than the location associated with *conceptual* color. Conceptual, or prototypical, color is a quite different kind of conceptual feature to category membership—a piano is a musical instrument whatever kind of piano it is, but what is its prototypical color? Black? Brown? White? And are apples green or red? And South American tree frogs can be any number of colors, ranging from yellow through red to blue. Further (around the UK) water is rarely *blue*. It is thus unclear what manner of abstraction results from the experience of the different colors of different things (even plants and leaves). Indeed, color appears to be largely context-dependent in a way that category membership is not (which is why, for the majority of their lifetimes, bananas are not yellow, and yet they are still fruit). We suspect that this distinction between category membership and prototypical color is the cause of the behavioral distinction between the corresponding competitor effects.

There is one fact about these data that is particularly noteworthy: Even in the presence of the target object, language-mediated eye movements are directed to objects that share certain characteristics, but not all, with the target specification determined by the unfolding word. That is, the eyes are directed *spuriously* to objects that are not intended by the speaker to attract the hearer's attention. Intriguingly, these spurious shifts in attention occur regardless of conceptual mismatch (as in the visual shape competitor and color competitor studies) or form mismatch (as in the conceptual competitor study). We return to the implications of such spurious eye movements after the next section, in which we use the competitor effects we have identified above (and specifically visual form competitor effects) to study the perceptual and attentional consequences of lexical ambiguity.

### 11.5 COMPETITOR EFFECTS AND THE ONLINE PROCESSING OF LEXICALLY AMBIGUOUS WORDS

Perceptual and conceptual competitor effects are a useful indication of the functioning of the cognitive processing system. But they can also be used as a tool to explore issues in language processing. The final study we shall describe uses competitor effects to explore issues in the processing of lexical (semantic) ambiguity.

**200** THE ON-LINE STUDY OF SENTENCE COMPREHENSION*11.5.1 Experiment 5: Semantically Ambiguous Words*

The indeterminacy of meaning in our environment is a particularly pervasive problem that faces the cognitive system. The interpretation of a word can vary greatly from context to context, and this is particularly true in the case of homonyms for which the alternative interpretations may be completely unrelated as in *The animal rights activist will locate the animals and then unlock the pen* and *The secretary will write the letter and then put down the pen*. Here, it is simply an accident of the phonology that the two tokens of *pen* can have such differing meanings. Research on lexical ambiguity has centered on the question of how we come to select and integrate into the context the appropriate meaning of an ambiguous word. Duffy, Morris, and Rayner (1988), for example, provided evidence that best supports models in which more than one meaning of an ambiguous word is initially activated, with the degree of activation being influenced by the fit with prior context and by the relative frequencies of the alternative meanings of the ambiguous word. Their data provided evidence against multiple access models that assume that context has no effect on lexical access and evidence against selective access models that assume that sufficiently strong contexts will lead to selective access of only the contextually appropriate meaning. Given the sensitivity of language-mediated eye movements to lexical access (e.g., Allopenna, Magnuson, & Tanenhaus, 1998) and contextual integration (e.g., Kamide, Altmann, & Haywood, 2003), we conducted a series of studies identical in methodology to the competitor studies described above but in which the target words were lexically ambiguous (thus, instead of *spinach*, or *piano*, we used words like *pen* and *diamond*). Specifically, we were interested in the temporal dynamics with which objects in the visual scene that were related to the different meanings of the ambiguous target word would be fixated.

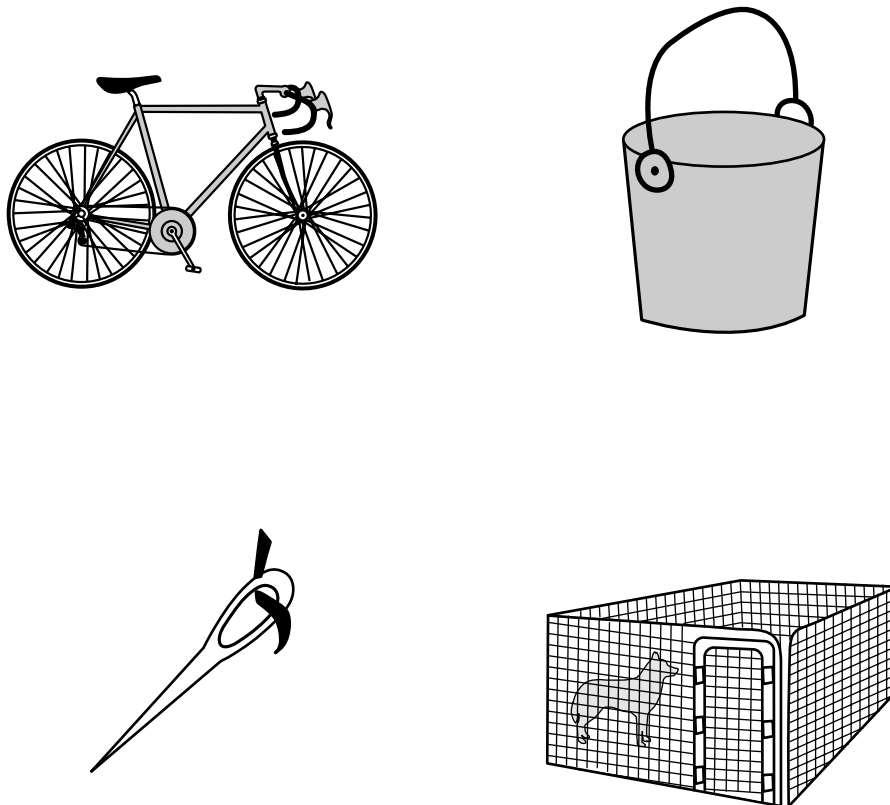
We selected 15 ambiguous words that are highly polarized with respect to the relative frequencies of their alternative meanings. Thus, for the word *pen*, the writing implement meaning is considerably more frequent than the enclosure meaning. We carried out a single word association task to determine the relative frequencies of these meanings. There were three experimental conditions in the first study: a neutral context condition, a biasing context condition, and a visual competitor condition. The neutral context did not bias one meaning or another of the target word: *First, the man got ready quickly, but then he checked the pen and suspected that it was damaged*. The sentences in the biasing and visual competitor conditions were identical but were designed to bias toward one meaning rather than another: *First, the welder locked up carefully, but then he checked the pen and suspected that it was damaged*. The sentences were identical to the neutral condition except for a phrase that biased towards the subordinate meaning of the ambiguous word (*welder locked up carefully*) which replaced the neutral phrase in the neutral condition (*man got ready quickly*). There has been a great deal of controversy about the classification of the strength of linguistic contexts in lexical ambiguity experiments (cf. Tanenhaus & Lucas, 1987). However, there has been little consensus about what constitutes a strong linguistic context. And although we used linguistic contexts that we judged would bias the subordinate meaning of ambiguous words, the evidence for the efficacy of our bias manipulation lies

EVIDENCE FROM HEAD-MOUNTED EYETRACKING **201**

in the empirical data—that is, whether or not the bias increases the likelihood of fixations on the subordinate object (the enclosure) at the onset of the target word *pen* (cf. our use of biasing contexts in the visual competitor study described in the preceding text).

Of primary interest in this study was the presence or absence of visual referents for the alternative meanings of the lexically ambiguous target word and the interaction of eye movements with the sentential context. Each visual stimulus in the neutral and biasing condition portrayed four objects: One corresponded to the dominant meaning of the target word (e.g., a writing pen), one corresponded to the subordinate meaning (e.g., an appropriate enclosure), and two depicted objects from different semantic/conceptual categories. The names of all the objects in the scene were frequency-matched according to the CELEX database. In the visual competitor condition, the object corresponding to the dominant meaning (e.g., *pen*—writing implement) was replaced with a visual competitor of a similar visual form (a needle; see Figure 11.6).

We shall refer to the object corresponding to the dominating meaning as the *dominant object*, and similarly for the subordinate object. The procedure we used



**Figure 11.6** An example stimulus of the visual competitor condition (the visual form competitor *needle* is depicted instead of the dominant meaning *pen*—writing implement).

**202** THE ON-LINE STUDY OF SENTENCE COMPREHENSION

was the same as in the previous experiments. We shall first report the percentage of trials on which each object was already being fixated at the onset of the target word. In the neutral condition, there were no significant differences in fixations to the dominant object, the subordinate object, or the distractor objects (24% on average for each one). Thus, the context was indeed neutral with respect to the likelihood of the dominant or subordinate objects being fixated. In the condition that favored the subordinate meaning, there were no differences in fixation likelihood between the distractors and the dominant object (17% each). However, there was a large increase in the likelihood that the subordinate object was being fixated by target onset (46%). Thus, our biasing contexts, which were designed to bias towards the subordinate meaning of the target word, appear to have been effective.<sup>5</sup> In the visual competitor condition (in which the dominant object was replaced by a visually similar competitor), there was no difference in fixation likelihood at target onset between the visual competitor and the distractors (17% each), but there was again a larger likelihood of fixation on the subordinate object (48%). We turn now to the incidence of saccadic movements during the target word itself (mean duration: 400 ms). In the neutral condition, there were marginally more saccades launched towards the dominant object than toward the subordinate object (26% vs. 17%—statistically significant by subjects but not by items). Both of these attracted more saccades than did the distractors (8% each). In the biasing condition, there were no more saccades launched toward the subordinate object (16%) than towards the dominant object (16%). Again, the distractors attracted fewer saccades (8% each). To the extent that we can interpret the neutral context condition as reflecting the standard dominance effect (with the dominant meaning being more strongly activated than the subordinate meaning), the biasing context suggests that looks toward the dominant object have been suppressed relative to the neutral condition (26% vs. 16%, a statistically significant difference, albeit at  $p = .05$  by items). Nonetheless, the dominant object *was* looked at more often than either of the distractors. Of interest, therefore, is what happened in the visual competitor condition where the dominant object was replaced by a visual competitor (in this example, a needle). In the event, there were more looks toward the visual competitor than to either of the distractors (21% vs. 11%—significant by subjects, and approaching significance by items:  $p < .1$ ), with looks to the subordinate object nonsignificantly lower than looks to the visual competitor (17%). The patterns of saccadic eye movements during the target word itself are mirrored more robustly, statistically speaking, in the data on the likelihood of each object being fixated at target word *offset*. In the neutral condition the dominant object was being fixated at this point on 43% of trials, compared to 29% for the subordinate object and 10% for each of the distractors. In the biasing condition, the corresponding figures were 27%, 46%, and 9%, respectively, and in the visual competitor condition, 26%, 50%, and 10%, respectively.

To summarize: We have converging evidence from the patterns of both fixation and saccades that in the neutral context the dominant object attracts more looks than the subordinate, which attracts more looks than the distractors. In the biasing context, the pattern reverses, with more fixations on the subordinate object than on the dominant one. Importantly, this reflects only the fact that there were more fixations to begin with on the subordinate object. In terms of *switched* attention, the dominant



EVIDENCE FROM HEAD-MOUNTED EYETRACKING **203**

and subordinate object attracted equal numbers of saccadic movements, suggesting that the effect of the biasing context is to reduce the otherwise large number of movements toward the dominant object that would occur in the absence of a bias. Of course, one could argue that this pattern of shifting attention is in part an artifact of the fact that both meanings of the target word are represented in the visual scene. Perhaps, in the biasing context, only the subordinate meaning of the ambiguous word would ordinarily be active, but the physical presence in the visual scene of an object corresponding to the dominant meaning causes that meaning to be active also. The visual competitor data address this issue; in this case, there was no dominant object, just a visual competitor. And in that condition, the data patterned almost identically with the case where the dominant meaning *was* represented (the biasing condition). Thus, even in the absence of an object corresponding to the dominant meaning we find strong evidence that conceptual information associated with the dominant meaning, regarding physical form, was accessed during the target word *pen*. Thus, and despite the linguistic and visual context biasing the subordinate meaning, we found compelling evidence of the activation of the dominant meaning of the ambiguous word.

This last finding is open to alternative interpretations. On the one hand, the dominant meaning of *pen* may indeed have been activated in the context that biased the subordinate meaning (as predicted by multiple access models and hybrid models of lexical ambiguity resolution), and this activated representation may then have mediated eye movements toward the object that shared with this representation certain visual form features. On the other hand, the activation of the dominant meaning in the biasing context may have been due in part to the existence in the scene of an object with those form features. This may have activated concepts sharing those features (cf. our earlier account of competitor effects), including the concept associated with the dominant meaning of the ambiguous word (in other words, the activation may have originated through spreading activation from form information portrayed within the visual scene). The current data do not distinguish between these two cases.

Finally, we repeated the biased condition, in which the sentential context originally favored the subordinate meaning of the target word, and in which the visual scene portrayed objects depicting both the dominant and subordinate meanings of *pen*. In the new version of this condition, we modified the sentential contexts to bias the *dominant* meaning. Would we still find increased looks toward both the dominant *and* subordinate objects relative to the distractors? There were indeed significantly more saccades during *pen* to both the dominant and subordinate objects (20% and 14%, respectively) than to the unrelated distractors (7% each). The difference in looks to the dominant and subordinate objects was not statistically reliable.

To summarize: The main finding that goes beyond previous studies using other methodologies is that the presence of a clear and contextually appropriate referent for one meaning of the lexically ambiguous item does not prevent eye movements to a potential referent for the other meaning. This is true irrespective of whether the context biases the subordinate meaning or the dominant meaning. Even more significantly, the presence of such a clear and appropriate referent does not even prevent eye movements to other objects that are *unrelated* conceptually to the alternative

## 204 THE ON-LINE STUDY OF SENTENCE COMPREHENSION

meaning (although they must share some features with it; in these cases, it was visual form). These data confirm our previous findings that the activation of meaning has behavioral consequences that go beyond attending only to the intended referent; these consequences include fast eye movements in response to an individual word toward objects that, although related in visual form, color, or conceptual category, are clearly not the intended referent for that word.

### 11.6 GENERAL DISCUSSION AND CONCLUSION

Our research with regard to lexically ambiguous words provides an important addendum to the literature to date. In some respects, our study with a visual competitor in place of the object corresponding to the dominant meaning of an ambiguous word is simply a visual-world analog of a standard cross-modal priming study. In such a study, the target word would be presented in the auditory modality, and reaction times to a related word presented subsequently in the visual modality would be measured at test. In our studies, we also presented the target word in the auditory modality, although instead of subsequently monitoring reaction times to a related word, we monitored concurrent eye movements to a related object (related through visual form, although we anticipate that we could just as well have used category competitors as in our earlier competitor studies). We observed that the reaction to that object was quantitatively different to the reaction toward unrelated objects. In some respects, therefore, we have simply “ported” one paradigm onto another. And what we have shown is that conceptual representations corresponding to the contextually less preferred meaning are nonetheless activated and nonetheless mediate visual attention about the concurrent visual scene. Importantly, such activation happens even when there is, quite unambiguously, a uniquely identifiable visual referent that corresponds to the intended (contextually biased) meaning.

This last observation is perhaps the most significant. These data, coupled with the competitor effects we described earlier, suggest that even when we know what it is in the concurrent environment that is being referred to, we cannot help but attend to other objects that are related. It is this observation that is in some respects the most challenging. It suggests that, during everyday conversation, there are pressures driving our visual attention spuriously toward objects in the environment that are not the intended ones to which we should be attending. The situation is made all the more dramatic when we consider data by Allopenna and colleagues showing that as a word such as *candy* unfolds, our attention is driven not just towards *candy*, but towards *candles* also (a “cohort competitor” effect; Allopenna et al., 1998). Our data go further in showing that as those mental representations unfold, our attention is driven also to objects that are related in various different ways to the target concept (and we have only just begun to investigate the dimensions of relatedness on which such effects might occur). And when a word has more than one meaning, each of those meanings exerts an influence on where the visual system should attend next, somewhat independently of the context, and certainly independently of whether or not whatever is being attended to matches that meaning on all the relevant dimensions. One challenge that these data present is that there must be some mechanism

EVIDENCE FROM HEAD-MOUNTED EYETRACKING **205**

that prevents the explosion of spurious shifts in attention that would occur if these influences went unchecked during normal conversational dialogue. In the absence of further data, we can only conjecture that such a mechanism must exist. Further research will be required to understand the limits on the effects we have identified here.

Eye movements to perceptual and conceptual competitors are fast, but are they automatic? This largely depends on our definition of automaticity. We do not have anything other than anecdotal reports from the participants themselves that they were unaware of any conscious control over their eye movements. The speed with which these effects manifest themselves (the target words averaged approximately 400 ms in duration) suggests that the eye movements we observed were not under volitional control (at least not in the sense that a conscious decision was made to move the eyes in a particular direction). Furthermore, we also obtained these results when only a small percentage of trials investigated a particular competitor relationship (less than 15%). Nor are our results an artifact of uncontrolled differences between different auditory targets or different visual objects. In one study we observed for the same acoustic target word *lettuce* more looks to a green jacket than toward blue trousers. But we also observed in this study fewer looks to the green jacket after *lemon*. Thus, whether we monitor for the same acoustic target looks toward different objects or, for the same object, looks engendered by different acoustic targets, the pattern is the same.

Traditional methods in experimental psycholinguistics have failed to highlight the behavioral consequences of a conceptual system whose concepts are not indivisible wholes. The major consequence we have highlighted in this chapter is that the conceptual parts that make up the whole can, during spoken word recognition, exert independent influences on how the cognitive system attends to the external world. One of the basic tenets of embodied cognition is that cognition is rooted in the same representational substrate that supports interaction with the external world. A direct consequence of this approach to cognition is that language, as a component part of cognition, must be studied in the context of the interactions it causes between the hearer and the world. One such interaction comprises the manner in which attention is directed, by language, around that world. We believe that, for too long, experimental psycholinguistics has remained modular, encapsulated, and theoretically autonomous with respect to the cognitive processes that serve our interactions with the world around us. Research into the relationship between language and vision attempts to place psycholinguistics at the center of cognition and, conversely, cognition at the center of psycholinguistics.

PE: Is this  
OK?

**AUTHOR NOTE/ACKNOWLEDGMENTS**

Correspondence should be sent to Falk Huettig, Department of Psychology, University of York, York, YO10 5DD, England, UK. Email: F.Huettig@psych.york.ac.uk. FH was supported by a University of York doctoral studentship. The research was also made possible with support from the Medical Research Council (G0000224 awarded to GA). We thank Gareth Gaskell, Graham Hitch, Ken McRae, and Mike

## 206 THE ON-LINE STUDY OF SENTENCE COMPREHENSION

Tanenhaus for constructive discussion of this research, and Chuck Clifton and Manuel Carreiras for their comments on an earlier version of this chapter.

### References

- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX lexical database [CD-ROM]. Philadelphia Linguistic Data Consortium, University of Pennsylvania.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Science*, 7, 84–91.
- Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 80.
- Calvin, W., & Bickerton, D. (2000). *Lingua ex machine: Reconciling Darwin and Chomsky with the human brain*. Cambridge, MA: MIT Press.
- Cree, G. S., & McRae, K. (2003, June). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*.
- Dahan, D., & Tanenhaus, M. (2002). *Activation of conceptual representations during spoken-word recognition*. Poster presented at the 43rd Annual Meeting of the Psychonomics Society, Kansas.
- Dahan, D., & Tanenhaus, M. K. (submitted). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints.
- Dell'Aqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition*, 73, B1–B15.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429–446.
- Glenberg, A. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1–55.
- Henderson, J. M., & Ferreira, F. (Eds.). (2003). *The interface of language, vision, and action*. Hove, UK: Psychology Press.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). The time course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.
- Kamide, Y., Scheepers, C., Altmann, G., & Crocker, M. (2002). *Integration of semantic and syntactic information in predictive processing*. Paper presented at the 15th CUNY conference, New York.
- McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 558–572.
- Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 863–883.

AU: Is there any additional information such as vols. or page number?

AU: Provide year of last draft as per APA style. Published yet?

AU: In which city was this held?

AU: Publisher location provided correct?

EVIDENCE FROM HEAD-MOUNTED EYETRACKING **207**

- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. From <http://www.usf.edu/Free-Association/>
- O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: the world as an outside memory. *Canadian Journal of Psychology*, *46*, 461–488.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardised set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.
- Tanenhaus, M. K., & Lucas, M. M. (1987). Context effects in lexical processing. *Cognition*, *25*, 213–234.
- Yee, E., & Sedivy, J. (2001). *Using eye movements to track the spread of semantic activation during spoken word recognition*. Paper presented to the 13th Annual CUNY Sentence Processing Conference, Philadelphia.

**Notes**

1. However, and in line with other recent work using the visual-world paradigm (see Henderson & Ferreira, 2003, for recent reviews), we, in fact, defined this region as starting 200 ms *after* the onset of the target word and ending 200 ms after its offset. This is to take into account the time it takes to program and initiate a saccadic movement (see Altmann & Kamide, in press, for an empirical investigation into such launch times, and Dahan & Tanenhaus, (submitted), for compelling evidence that the 200 ms figure is an accurate estimate).
2. Henceforth, all reported differences in subsequent analyses were significant by both subjects and items unless otherwise described.
3. In collaboration with Ken McRae, we computed the "conceptual distance" between each target and its conceptual competitor (see Cree & McRae, 2003, for a discussion of the relevant semantic feature norms). Of the 30 target-competitor pairs in our study, we could compute such distance scores for 24 of the pairs (that is, existing semantic feature norms existed for both members of the pair). We eliminated a further two pairs whose members were visually very similar. We then correlated the remaining distance scores against the proportion of saccades launched during the target word towards the competitor in the competitor-only condition. The resulting correlation was highly significant ( $r = 0.6$ ), and we are thus confident that our data do, indeed, reflect effects mediated at the level of semantic features.
4. It is the change in activation state that causes the shift in attention, because, although the representation corresponding to the trumpet receives some priming from *piano*, we can suppose that it becomes neither as active as the representation corresponding to whatever is currently being fixated, nor as active as the representation corresponding to *piano*. Thus, we conjecture that it is not *degree* of activation but *change* in activation that drives shifts in attention. However, a full treatment of what drives attention is beyond the remit of this research. Nonetheless, to fully understand how it is that language can mediate visual attention will require an understanding also of attentional control.
5. See also the literature on prediction effects in sentence comprehension (e.g., Altmann & Kamide, 1999; Kamide, Scheepers, Altmann, & Crocker, 2002; McDonald & Shillcock, chapter 5, this volume).

