

Eye Movements to Pictures Reveal Transient Semantic Activation During Spoken Word Recognition

Eiling Yee and Julie C. Sedivy
Brown University

Two experiments explore the activation of semantic information during spoken word recognition. Experiment 1 shows that as the name of an object unfolds (e.g., *lock*), eye movements are drawn to pictorial representations of both the named object and semantically related objects (e.g., *key*). Experiment 2 shows that objects semantically related to an uttered word's onset competitors become active enough to draw visual attention (e.g., if the uttered word is *logs*, participants fixate on *key* because of partial activation of *lock*), despite that the onset competitor itself is not present in the visual display. Together, these experiments provide detailed information about the activation of semantic information associated with a spoken word and its phonological competitors and demonstrate that transient semantic activation is sufficient to impact visual attention.

Keywords: spoken word recognition, eye movements, cohort competition, semantic activation, lexical access

Human beings typically take for granted the ability to easily understand what is said. Successful spoken word recognition, however, is more complicated than it seems. People must accurately select both the form and the meaning of the uttered word from among the thousands of candidates in their mental lexicons, despite the fact that the speech heard at a given moment is usually consistent with any one of a large number of words (e.g., */la/* could be the start of *law*, *lock*, *logs*, *lost*, *lobster*).

Most models of spoken word recognition focus on describing how we access the appropriate lexical form from among all of the potentially matching candidates (e.g., Marslen-Wilson, & Welsh's [1978] cohort model; McClelland & Elman's [1986] TRACE model; Norris's [1994] shortlist model; and Luce & Pisoni's [1998] neighborhood activation model). Although differing on the details, these models all agree that as a given spoken word unfolds, words that start with the same sounds become partially active (i.e., hearing the */la/* of *lock* triggers activation of phonologically related words like *logs*).

In contrast to their emphasis on form, few models of spoken word recognition explicitly address access to meaning (but cf. Gaskell & Marslen-Wilson, 1997 and McNellis & Blumstein, 2001). Nevertheless, it is generally assumed that when a word's form is accessed, its meaning is also automatically accessed (but cf. Connine, Titone, Deelman, & Blasko, 1997). If it is true that phonologically related words are partially activated when a word is heard and also that form and meaning are activated together, this leaves open the possibility that people access the meanings of multiple unintended candidates before finally settling on the intended one. Although linguistic and extralinguistic context may often help to reduce this risk, there are times when context does little if anything to narrow down the choices. Clearly, all this spurious semantic activation would have the potential to cause listeners a great deal of confusion. Yet people are very rarely aware of considering candidates that turn out to be incorrect. This absence of conscious confusion may appear to indicate that semantic representations of unintended candidates do not become active. However, there is evidence that as people hear a given spoken word they do, in fact, temporarily access semantic information about words with the same onset (henceforth *onset competitors*). Most of the existing evidence comes from experiments that use the cross-modal semantic priming paradigm.¹ In these experiments, a written target word is presented before the offset of an auditory prime—just before the prime would have become unambiguous (e.g., if the prime were *lock* or *logs*, the target would be presented at the end of */la/*). Participants make faster lexical decisions when the targets are related to possible continuations of the prime's onset (e.g., *key* or *wood*) than when targets are unrelated to the prime. Because words related to both the prime and its

Eiling Yee and Julie C. Sedivy, Department of Cognitive and Linguistic Sciences, Brown University.

This research was partially funded by National Institutes of Health Grant R01 MH62566-01A1 and by a Jacob K. Javits Fellowship awarded to Eiling Yee. We thank Paul Allopenna, Sheila Blumstein, William Heindel, Jesse Hochstadt, James Morgan, and Katherine White for their extremely helpful comments and contributions to this project. We also thank Michelle Engle and Anjula Joshi for assistance with data collection, and Andrew Duchon for essential assistance with data processing. Portions of this research were presented at the 14th Annual CUNY Conference on Human Sentence Processing, March, 2001 and at the 42nd Annual Meeting of the Psychonomic Society, November 2001.

Correspondence concerning this article should be addressed to Eiling Yee, who is now at the University of Pennsylvania, Department of Psychology, 3720 Walnut Street, Philadelphia, PA 19104-6241. E-mail: eiling@psych.upenn.edu

¹ Although gating and phonological priming (also known as *form priming*) tasks have been used to investigate whether a word's phonological competitors become active, results from these paradigms cannot indicate whether a word's meaning has been activated.

onset competitors appear to become active, these results suggest that initially people do activate both the form and the meaning of multiple potential candidates (e.g., Marslen-Wilson, 1987; Moss, McCormick, & Tyler, 1997; Zwitserlood, 1989).

Yet the fact that people are not consciously aware of activating the meanings of unintended candidates suggests that these candidates may not become active enough to affect behavior in non-laboratory settings. That is, the effects observed in cross-modal semantic priming studies may not generalize to more natural settings. This is a reason for concern because the task in this paradigm is somewhat unnatural. For example, presenting the target before the prime's offset could lead participants to place unnaturally heavy weight on the prime's onset by interrupting the processing of the prime. Furthermore, participants are required to make overt judgments about targets that are paired with semantically related primes, making the relationship between the two words quite conspicuous, which can result in task-specific effects. For example, it has been found that as the proportion of related trials increases, priming effects increase (Tweedy, Lapinski, & Schaneveldt, 1977). It would therefore be useful to gauge semantic activation of related lexical items in a task that does not make this relationship salient, and ideally, one in which no response to the related item itself is required.

Setting aside issues of ecological validity, cross-modal semantic priming (like all semantic priming paradigms), has a significant limitation in that it provides a very indirect measure of the prime word's activation, showing only the extent to which the activation of the prime word facilitates responses to the target word. It does not, unfortunately, directly reflect the activation of the prime word itself. Because priming is a composite measure, combining the prime's activation with potential effects of that activation on responses to a target, priming effects tend to be fairly variable across tasks. For example, they have been shown to vary depending upon a number of factors, including whether the prime is represented pictorially or verbally, and the nature of the required response to the target (e.g., naming, lexical decision, semantic categorization). Furthermore, because the method only provides information about a specifically sampled point in time, it is necessary to conduct multiple experiments, with varying interstimulus intervals, in order to assess how a word's activation changes over time.

It has been suggested that the relatively recent method of monitoring eye movements to a visual display during spoken word processing has the potential to address the above concerns and provides a more direct reflection of the activation of a word's representation. In this "visual world" paradigm, participants' eye movements are recorded as they follow spoken instructions to point to or manipulate an object that is part of a visual display in front of them. It has been found that there is a close time locking between an unfolding referring expression and eye movements to a potential referent, making eye movements a valuable indicator of which lexical candidates participants are considering (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). A significant benefit of this technique is that eye movements can be measured without disrupting the spoken input or requiring participants to make a metalinguistic judgment such as lexical decision. This allows participants to engage in tasks that are more naturalistic than those used in standard psycholinguistic paradigms. Most

important, eye movements provide a continuous measure of processing as participants listen to language input. Thus they supply information about word recognition as the word unfolds.

There has been some encouraging convergence between results obtained with this method and results from more traditional tasks. For example, several eye-tracking studies have shown that eye movements tend to be drawn to onset competitors of a given spoken word (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Dahan, Magnuson, & Tanenhaus, 2001; Tanenhaus et al., 1995). In these studies, participants are presented with a four-picture display and asked to "pick up" (i.e., move with a computer mouse) one of the objects (the target). It has been found that if the name of one of the objects is an onset competitor of the target word, participants are initially more likely to fixate on this onset competitor rather than on objects with phonologically unrelated names. For example, if asked to "Pick up the *beaker*," participants are more likely to fixate on the onset competitor *beetle* than on objects with phonologically unrelated names. Intriguingly, Allopenna et al. (1998) also showed that as a word unfolds, the likelihood that a participant will fixate on the corresponding picture—and also on its onset competitors—closely matches the word's lexical activation as predicted by simulations that use the TRACE model of spoken word recognition. This correspondence suggests that in active tasks such as this one, participants' fixations are tightly linked to lexical activation.

More recently, Huettig and Altmann (2005) have measured eye movements in a passive listening task to investigate the activation of semantic information. In this study, participants were instructed to scan visual displays accompanying sentences such as *Eventually the man agreed hesitantly, but then he looked at the piano and appreciated that it was beautiful*. Participants were not required to carry out any task or to produce a response. Upon hearing the word *piano*, they showed a greater tendency to fixate on a picture of a semantically related object, such as a trumpet, than on unrelated distractor objects in the display. This effect occurred beginning at approximately 300–400 ms after the onset of the target word and was interpreted as evidence that as a word is heard, visual attention can be rapidly directed to semantically related words. Patterns of eye movements, therefore, have been shown to reflect the partial activation of words related to target words along both phonological and semantic dimensions.

However, the visual world paradigm introduces its own set of challenges and potential limitations, arising from the fact that the dependent measure requires the simultaneous or preceding presentation of a visual stimulus corresponding to the verbal one. This has several implications for the application of the visual world method to the study of lexical activation. First is the possibility that exposure to a picture results in the activation of its corresponding name, independently of any verbal stimulus. Although priming from pictures to words does not appear to be as robust as within-modality priming (e.g., from pictures to pictures or from words to words), under some circumstances responses to a word target are speeded when the word is preceded by a corresponding or related picture. For example, Vanderwart (1984) found that pictures facilitated responses to related words in a lexical decision task, though the priming effect was smaller than that obtained when word primes were used. Similar results were found by Carr, McCauley, and Parmelee (1982) through the use of a naming task. Other studies, however, have failed to find priming effects from

pictures to words. Durso and Johnson (1979) reported that although word primes facilitated the naming of their corresponding pictures, picture primes did not facilitate the naming of corresponding words. Scarborough, Gerard, and Cortese (1979) found a similar pattern in a lexical decision task. More recently, Dell'Acqua and Grainger (1999) reported that subliminally presented pictures facilitated responses to related and corresponding words in a semantic categorization task but not in a naming task. Bajo (1988) found robust priming effects for pictures as targets, but weaker picture-to-word priming, and noted that priming effects with word targets were bolstered when the task encouraged semantic processing (categorization vs. naming), when participants were told that the relation between the prime and target was important, and when a blocked design was used, in which trials were grouped according to whether primes and targets were pictures versus words rather than being intermixed.

On the basis of priming experiments, then, it is possible that the presentation of a picture in a visual world task automatically primes its name or related names. It should be noted that this possibility is amplified by the fact that visual world experiments generally do encourage semantic processing; for example, they frequently use a task in which participants are required to identify referents corresponding to spoken stimuli while carrying out a specific instruction. Indeed, this referential aspect of the paradigm increases its ecological validity. However, at the same time, it introduces the possibility of influence from the visual domain on the processing of the speech stimulus. It is important to understand the nature of this influence if the visual world paradigm is to be used to make inferences about language processing.

A second source of possible influence from the visual sphere pertains to circumscribing the referential domain, particularly in tasks in which linguistic stimuli are explicitly linked to objects in the visual display. In such tasks, reference resolution requires both activating the referent's lexical representation and referentially mapping that representation onto the immediate visual environment. It has been shown that referential commitments can be affected by numerous high level factors, including perspective-taking (Hanna, Tanenhaus, & Trueswell, 2003; Nadig & Sedivy, 2002), object affordances (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2001), pragmatic inferences (Sedivy, Tanenhaus, Chambers, & Carlson, 1999), and the structural combination of the words in the sentence and knowledge of the visual context (Tanenhaus et al., 1995). In all cases, inferences about possible or likely referents provide strong constraints on how spoken stimuli are mapped onto the visual scene.

For example, the visual world paradigm has been productively used to study temporarily ambiguous sentence structures such as *Put the apple on the towel into the box* (e.g., Spivey, Tanenhaus, Eberhard, & Sedivy, 2002), in which the phrase *on the towel* is temporarily consistent with either an interpretation in which it is describing the current location of the apple (as in *Put the apple that's on the towel. . .*) or as the location where the participant is meant to place the apple. If the display includes two towels, each relevant for only one of the possible interpretations of the phrase *on the towel*, eye movements are not equally distributed over the two towels. Instead, eye movement patterns differ depending on the structural analysis assigned to the phrase. This indicates that activation of the lexical item *towel* is not sufficient to draw fixations to both towels; rather, the eye movements reflect how

participants are assigning reference on the basis of the structural combination of the words in the sentence and knowledge of the visual context. Thus, it is clear that in referential tasks, eye movements measure more than the activation of a lexical representation. They also reflect the mapping of that representation onto the visual scene. Non-lexical factors are important for this latter process, influencing which objects in the display are considered likely or possible referents.

Since it constrains potential referents to the objects in the display, the visual environment is clearly a factor that constrains referential mapping. It is possible, however, that the visual environment also constrains *lexical processing* to the objects that are in the display. That is, the display may serve to delimit a "closed set" of representations consistent with the spoken stimuli, and lexical processing may occur only within this "closed-set". If true, patterns of eye movements may not be driven by activation levels within the lexical system as a whole but rather reflect processing of only those lexical items that may be mapped onto possible referents (i.e., objects in the display). Therefore, it is of considerable methodological and theoretical interest to determine how lexical processing is affected by referential constraints. Fortunately, the visual world paradigm provides information about more than just the way that reference is ultimately resolved. It also supplies information about time course and partial activation, providing insights into the processing that occurs *as* reference is being resolved. These sources of information make it possible to assess whether lexical processing is limited to the objects in the visual environment.

Two recent studies have used time course information to explore whether lexical processing is limited to the closed set of objects in the display (Dahan et al., 2001b; Dahan et al., 2001a). In a visual world study of lexical processing, Dahan et al., 2001b investigated the effects of frequency on eye-movement patterns. Although frequency effects have long been documented in reaction time studies of lexical processing, they can be reduced or eliminated in closed-set tests in which all response alternatives are equally likely. Hence, frequency effects should only arise in the visual world paradigm if eye movements reflect general properties of lexical processing. Dahan et al. found that when the display contained two pictures whose names had overlapping onsets (e.g., *bed* and *bell*) participants were more likely to fixate on the picture associated with the higher frequency name during the initial portion of the word. Furthermore, when presented with displays containing a target referent and three phonologically unrelated distractors, the target was fixated on more rapidly when its name was a high-frequency word than when it was a low-frequency word. These results demonstrate that lexical processing is influenced by representations whose referents are not present in the display.

Additional evidence that the set of candidates that listeners consider is not limited to the names of the objects present in the display comes from Dahan et al. (2001a). In this study, participants were slower to fixate on a target picture if the onset of the target word was cross-spliced from an unpictured real-word competitor (e.g., the *ne* from *neck* cross-spliced into *net*) than they were if the onset was cross-spliced from a nonword (e.g., the *ne* from *nep* cross-spliced into *net*). Thus, participants were slower to fixate on the target when the inconsistent coarticulatory information came from a potential real-word competitor (despite that it was not

displayed) than they were when the inconsistent information came from a nonword. The results of both of these visual world studies indicate that the lexical forms of non-depicted items become partially active and can influence lexical processing. However, they do not address whether the semantic representations (in addition to the lexical forms) of non-depicted items are also activated.

Thus, while the visual world paradigm has potential limitations (i.e., exposure to a picture may result in the activation of its corresponding name independently of any verbal stimulus, and lexical-semantic processing may be limited to objects present in the display), it also has benefits over semantic priming paradigms (i.e., it is more natural, provides a more direct measure of lexical activation, and provides time course information). In this article we exploit these properties to explore the activation of semantic information during spoken word recognition. Experiment 1 investigates whether, in a task that requires identifying the visual referent of a spoken target word, eye movements are drawn to a semantically related object despite the absence of phonological overlap between the name of the target and the semantically related object. This experiment avoids some of the limitations inherent to traditional and cross-modal semantic priming paradigms because there is no task-related motivation to attend to the semantically related object and because time course data can be obtained. Experiment 1 contrasts with the Huettig and Altmann (2005) study in that it uses a task in which there is an explicit assumption that the spoken word is to be referentially linked to one of the pictures in the display. In the Huettig and Altmann study, there was no presumed link between the displays and the auditory input, and in fact, on a portion of the trials the displays did not contain a picture corresponding to the target word. Our task in Experiment 1 more closely resembles natural communication in that both speaker and listener assume a referential mapping between the uttered words and the visual environment to which the listener's attention has been directed.

However, the design of Experiment 1 leaves open the possibility that lexical-semantic processing is limited to a closed set of candidate referents rather than occurring in the context of the general lexicon. It also leaves open the possibility that any eye movements to a semantically related object are caused not only by activation of lexical items in response to spoken stimuli but also by the presence of two related objects in the display. In order to investigate these possibilities, Experiment 2 explores whether eye movements are drawn to an object semantically related to an uttered word's onset competitor—even though the competitor itself is absent from the display. By removing one of the semantically related objects from the display, Experiment 2 allows us to test whether lexical-semantic processing occurs within a closed-set of candidate referents and also eliminates any influence that the simple presence of related objects might have on lexical activation. A comparison of the data across the two experiments reveals the extent to which any semantic relatedness effects are driven by the visual availability of semantically related pictures. In addition, by measuring the activation of objects semantically related to onset competitors of the target word rather than to the target itself, Experiment 2 allows us to probe how the semantic relatedness effect evolves as the onset competitor diverges from the speech signal.

Experiment 1

Experiment 1 is intended to explore the time course over which an uttered word becomes semantically active by observing fixations on a related picture in the display. It could be argued that fixations on the picture corresponding to the target word itself show evidence that the word's semantic representation has been activated. After all, one might assume that in order to map a word onto its referent picture, the word's semantic representation must be active. However, there are reasons to be cautious when drawing inferences about semantic activation from eye movements to a picture of an object whose name matches the acoustic signal.

First, a recent study by Smith, Meiran, and Besner (2000) has suggested that it is not the case that a word's full semantic representation must be active in order to associate it with a picture. In their study of word-to-picture priming, form-based rather than semantic-based processing of the prime word was encouraged by requiring participants to perform a letter search task on the prime word. This type of task has been shown to interfere with semantic priming effects. In the Smith et al. (2000) study, when participants conducted a letter search on the word prime, there was no facilitation of responses to a picture semantically related to the word; however, facilitation effects were found for pictures depicting the word prime itself. These results suggest that a fairly direct associative link exists between words and corresponding pictorial representations: Mapping of words to corresponding pictures may occur even when semantic information is too weakly or incompletely activated to facilitate the processing of semantically related objects. This raises the possibility that direct word-to-referent mappings in a visual domain may not necessarily reflect full activation of semantic or conceptual information.

Second, it is possible that eye movements to a target can be driven by the match between the acoustic input and the name of the pictured object rather than by its meaning. Suppose that viewing a pictorial representation of an object automatically activates not only its semantic representation but also its phonological form (i.e., its name) and that once activated both the semantic and the phonological information become associated with the location of the picture. If true, eye movements to the location of a word's referent could commence even before the heard word itself has become active enough to be linked to its meaning. In other words, fixations to the location of a word's referent could be launched solely in response to the match between the phonological form of the heard word and the particular location in the display already associated with that form. Hence, in this scenario, acoustic input could trigger eye movements to the location of a picture even before that acoustic input activates the word's meaning.

Eye movements to objects semantically but not phonologically related to the target can be used to measure the semantic activation of the uttered word without being subject to the two concerns described above. If, when given the task of identifying a spoken target, participants are more likely to fixate on an object semantically (but not phonologically) related to a target than on unrelated objects, then this would provide evidence that eye movements reflect the activation of semantic representations rather than reflecting the match between acoustic input and phonological form alone. Perhaps more interestingly, such a finding would also demonstrate that words semantically related to the target become active enough for their pictorial representations to draw visual

attention, even when they are irrelevant to the task at hand. Although a great deal of previous work has shown that words semantically related to uttered words are activated relative to some baseline, the majority of these studies have required participants to name or make an overt judgment on the semantically related item itself.

Method

Participants. Thirty participants from the Brown University community were tested. All participants were native speakers of English and had normal or corrected-to-normal vision and reported no hearing deficits. They were paid a rate of \$7/hr for participating.

Apparatus. An SMI EyeLink I head-mounted eye tracker was used to monitor participants' eye movements. A camera imaged the participant's left eye at 250 Hz. Stimuli were presented with PsyScript (Bates & Oliveiro, 2003), a freely available language for scripting psychology experiments, on a 15-in. ELO touch-sensitive monitor.

Materials. We selected 24 pairs of semantically related stimuli. To minimize the possibility that participants would mistake the picture of the semantically related object for the target, only visually dissimilar target-related object pairs were selected. Object names were one to three syllables long (with an average of one and one half syllables). A full list of experimental items is given in Appendix A. For the displays, we selected color line drawings from a commercial clip art collection and from a collection of color line drawings (Rossion & Pourtois, 2001) based on the black and white Snodgrass picture library (Snodgrass & Vanderwart, 1980)

To ensure that the pictures in critical trials clearly represented what they were intended to represent, we conducted picture-name correspondence pretests. Participants who did not participate in either of the eye-tracking experiments were presented with each picture and a label (either its intended name or a randomly selected name) and were asked to judge whether they matched. To ensure a high degree of picture-name correspondence, at least 15 of the 16 participants had to agree that the intended name matched the picture. The few pictures that did not meet this criterion were replaced with new pictures. These new pictures were presented to at least 5 participants (not participating in the experiments) who were asked to name each picture. If more than 1 of the participants did not provide the intended name for a picture it was replaced with a new picture that was normed in the same way.

A female speaker (Eiling Yee), in a sound-treated room, read each target word in isolation with sentence-final intonation. Average target duration was 510 ms. The stimuli were recorded on a DAT tape and digitized at 20 kHz. Two lists, each 72 trials long, were created. Related word pairs appeared as target and related object on one list and as objects unrelated to the target on the other (Appendix A). Each participant was presented with only one list so that no participant saw or heard any object more than once.

Figure 1 shows a sample display. In the semantically related condition (12 of the 72 trials on each list), one of the objects in the display (e.g., *key*) was semantically related to the target (e.g., *lock*). We refer to this object as the *semantically related object*. (Except where specified, none of the objects in a display were semantically or phonologically related to any of the other objects in the display.) The other two objects were semantically and phonologically unrelated to both the target and the semantically related object. The name of one of these unrelated objects was matched for frequency with the semantically related object.² We refer to this object as the *related object's control*. The name of the other unrelated object (*the target's control*) was matched for frequency with the target.

The displays that were used in the semantically related condition in one list appeared in the control condition (12 of 72 trials) on the other list. This was accomplished by using one of the objects that had served as an unrelated object in the semantically related condition as the target (e.g., for the display in Figure 1, *deer* became the target in the control condition). Similarly, displays that were used in the control condition on one list

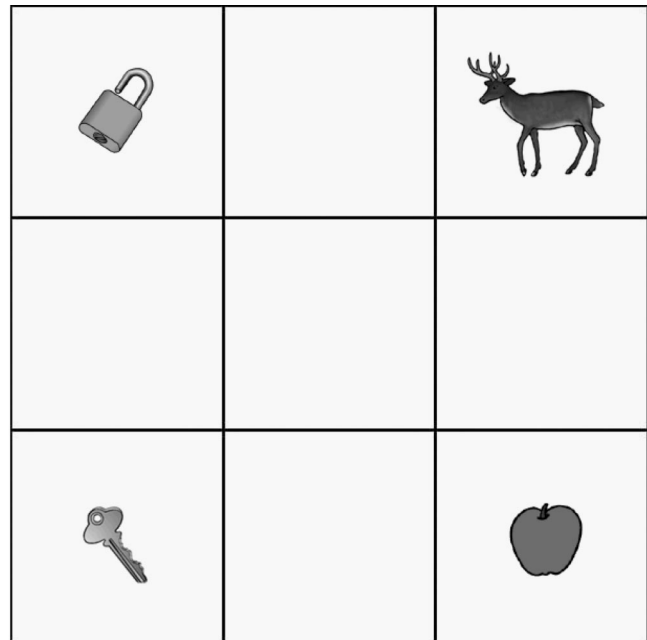


Figure 1. A sample display from Experiment 1. The target object, *lock*, is semantically related to one of the other objects in the display (the semantically related object, *key*). The other two objects are unrelated semantically and phonologically to the target and related object. This same display also appears (between participants) in the control condition, with the target's control (*deer*) serving as the target.

appeared in the semantically related condition on the other list. Thus, relatedness was manipulated within-subjects. The target in the control condition was the object that was frequency matched with the target in the semantically related condition (log frequency = 1.53 in control vs. 1.52 in experimental). Target durations and average number of syllables and duration were also similar (1.5 and 589 ms in control vs. 1.6 and 510 ms in experimental).

Because the same displays were used (between subjects) in both the semantically related and the control conditions, one of the nontarget objects in the control condition served as the related object in the semantically related condition. This made it possible to determine whether or not the images that served as related objects drew fixations regardless of their relationship to the target (e.g., because the pictures were more inherently interesting than the others in the display). Another benefit of this design was that in the control condition, although two of the objects in the display were semantically related to each other, neither one was related to the target. Therefore, if any participants noticed that some of the objects were related, they could not then predict over the course of the experiment that the target would be one of the related objects. Moreover, having two related objects in the display made it possible to determine whether the presence of a pair of semantically related objects draws participants' attention irrespective of the instructions. In filler trials (48 trials), no objects in the display were related in any way. Object positions, including the positional relationship between the target and the semantically related

² Each word's frequency count in the Brown corpus (Francis & Kucera, 1982), the Wall Street Journal corpus (Mitchell, Santorini & Marcinkiewicz, 1993), and the SWITCHBOARD corpus (Godfrey, Holliman & McDaniel, 1992), was obtained. For each word the three counts were summed; these logged sums were matched.

object, were balanced so that each object type was equally likely to appear in each corner of the display. Trial order was randomized for each participant.

Procedure. Participants were presented with a 3×3 array with four pictures on it, one in each corner (see Figure 1). Each cell in the array was approximately 2×2 in. Participants were seated at a comfortable distance (about 18 in.) from a touch-sensitive monitor, with the monitor at eye level. Therefore, each cell in the grid subtended about 6.4° of visual angle. (The eye tracker is accurate to less than 1° of visual angle.) A red square appeared in the center of the screen 1 s after the display appeared. Participants were instructed to touch the red square when it appeared. Touching the red square caused it to disappear and also triggered a sound file naming one of the objects in the display. The red square was included in the procedure to decrease the likelihood that participants would be fixating on one of the pictures at word onset. After the participant selected one of the pictures by touching it on the screen, the trial ended and the screen went blank. At this point the experimenter could either press a key to go on to the next trial or check the calibration before continuing. The experimenter continuously monitored the participants' performance and eye movements and suggested a break or validated the calibration as necessary. There were four practice trials.

Eye movements were recorded starting from when the array appeared on the screen and ending when the participant touched the screen to select a picture. Only fixations that were initiated after the onset of the target word were included in our analyses.³ We defined four regions, each corresponding to a corner cell in the array. The SMI software parses the eye-movement data into fixations, blinks, and saccades. We defined a *fixation* on a particular region as starting with the beginning of the saccade that moved into that region and ending with the beginning of the saccade that exited that region. (Therefore, any region-internal saccades that occurred in the interim were counted as part of a single fixation in that region.) In order for a fixation on a region to be counted it had to last at least 100 ms.

Results and Discussion

Four trials (0.6%) were not included in the analysis because the wrong picture was selected. Six percent of trials did not provide any data because there were no eye movements after the onset of the target word (most of these were trials in which the participant was already fixating on the picture of the target at the onset of the target word). For the remaining data, we computed the proportions (across trials) of fixations on each picture type (e.g., target, semantically related, unrelated) over time in 32-ms bins. Fixations anywhere inside the cell that contained a picture were counted as fixations on that picture. Fixation proportions more than 2.5 standard deviations from the mean were replaced with the mean of the remaining fixation proportions for that bin of that condition (3% by participants, 3% by items).⁴

Figure 2 plots the mean proportion of trials over time that contained a fixation on the target, on the semantically related object, and on the related object's control (from target onset to 1,000 ms after onset) in the semantically related condition. For the purpose of analyzing the data, we defined a *trial* as starting at 200 ms after target onset (because it takes a minimum of about 180 ms to initiate a saccade to a target in response to linguistic input when the specific target is not known ahead of time but when the possible locations of the target are known; Altmann & Kamide, [2004]) and ending at the point at which the probability of fixating on the target asymptoted. In these data, the end of the trial occurred at about 900 ms after target onset. Averaging fixations over the entire trial showed a greater proportion of fixations on the semantically related object's picture relative to its control, $t_1(29) = 6.9$,

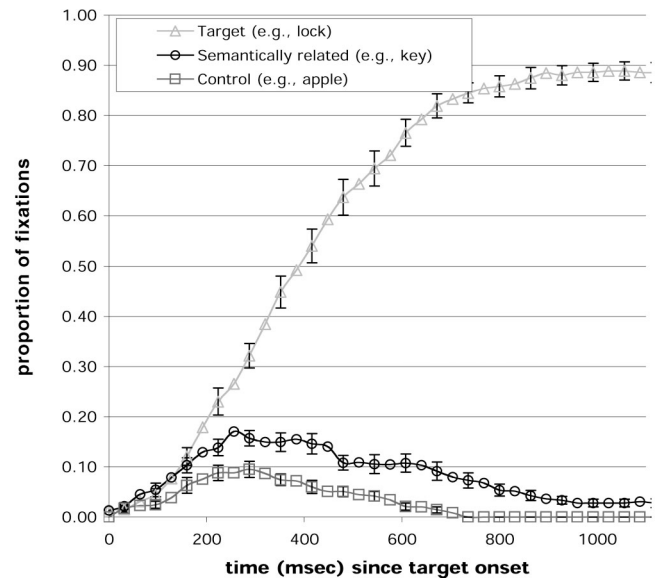


Figure 2. Experiment 1: Proportion of fixations over time on the target, the semantically related object, and the related object's control. Standard error bars are shown for every other data point.

$p < .01$; $t_2(23) = 5.2$, $p < .01$, for participants and items, respectively. Recall that the related object's control appeared in the same display as the related object and the target and was frequency matched with the related object. The semantic effect was also measured by comparing fixations on the related object when it appeared in the semantically related condition, with fixations on the same object in the control condition, where it was unrelated to the target: by participants, $t_1(29) = 7.4$, $p < .01$; by items, $t_2(23) = 12.4$, $p < .01$.

To obtain information about the time course of participants' fixations in the semantically related condition, we divided the trial into seven 100-ms windows (200–900 ms after target onset) and conducted separate planned comparisons on each window. (All differences reported below were reliable at $p < .05$ [one-tailed] for both participants and items.) First we sought to determine when the target started to become active by comparing fixations on the target picture with the average of fixations on the unrelated pictures. At the window beginning 200 ms after target onset, fixations on the target were significantly more likely than fixations on the unrelated pictures. This difference remained significant throughout the trial. Next, we established the time course of the semantically related object's activation by comparing fixations on the related object's picture with those on its control picture. At the window beginning 200 ms after the onset of the target, fixations on the semantically related object were more likely than were fixations on the related object's control. This difference also remained significant throughout the trial.

³ When fixations that were ongoing at the onset of the target word were included in the analyses reported in this article the patterns remained the same and the significance levels were largely unchanged.

⁴ When the analyses were repeated on untrimmed data the patterns remained the same and the significance levels were largely unchanged.

The results of Experiment 1 show that pictures of objects that are semantically but not phonologically related to the target draw more fixations than do pictures of unrelated objects. These fixations cannot be attributed to participants' simply matching the acoustic input with phonological form, suggesting that eye movements do in fact reflect the activation of semantic information. Fixations on semantically related objects began to increase in the window 200–300 ms after target onset, and they remained significantly above those on unrelated pictures until the end of the trial. The control condition ruled out two possible alternative explanations for the results. First, the same pictures were fixated on more frequently when they were semantically related to the target than when they were not related to the target, indicating that the pictures we used to represent the semantically related objects were not inherently more interesting than other pictures in the display. Second, when two pictures were related to each other but unrelated to the target, they did not draw more fixations than the unrelated picture. As a further test of whether objects drew disproportional visual attention merely because they were related to each other, we examined eye movements in the interval during which participants were exposed to the display prior to the acoustic input. On average, participants were exposed to the display for about 2 s prior to the acoustic input (1 s prior to the red square appearing, and then 955 ms [$SD = 347$ ms] before they touched the red square). We found that during this preword interval there was no difference between the proportion of fixations on the four objects, $F(3, 92) = .64, p = .59$. This finding indicates that participants' visual attention was not simply drawn to related pictures irrespective of the acoustic input. This is important because it shows that eye movements to related pictures were driven by the utterance of the target word.

However, it is necessary to rule out the possibility that despite our efforts to select only visually dissimilar pairs, participants fixated on related objects because they temporarily mistook them for the target. To determine whether mistaking the semantically related object for the target could account for the results, we conducted a visual similarity post-test. Our hypothesis was that if the results were due to visual confusion, then the visual similarity of related pairs (the target and the semantically related object, e.g., *lock* and *key*) should be higher than those of unrelated pairs (the target and the semantically related object's control, e.g., *lock* and *apple*).

We presented 24 participants (who had not participated in Experiment 1) with a written word and then with a picture. The instructions were as follows

You will see a word on the screen. Form a mental image of the object that the word refers to. Next you will see a picture. Rate the picture's shape according to how similar it is to the mental image you formed.

The word appeared on the screen for one second before the picture appeared, and both the word and the picture remained on the screen until the participant responded.⁵ Ratings were done on a 1–7 scale where 1 = *very different* and 7 = *very similar*. There were three lists and no participant saw any word or picture more than once. The presentation order was randomized for each participant. Similarity ratings for targets and related objects were quite low ($M = 2.4$), indicating that the selected pairs of objects were visually dissimilar. However, similarity ratings for targets and control pictures were even lower ($M = 1.5$), and the difference between the ratings for the related vs. the unrelated pairs was

statistically significant, $t_1(23) = 5.1, p < .01$, and $t_2(23) = 4.4, p < .01$ (one-tailed) for participants and items, respectively. To adjust for this disparity, we analyzed separately eight items that had equivalent ($M = 1.8$) average similarity ratings for the related object vs. the target and the control vs. the target. The pattern of results remained the same (Figure 3, left panel), with the semantically related object being fixated on significantly more often than the related object's control. Averaged over the trial, this result was significant by both participants, $t_1(29) = 3.5, p < .01$, and items, $t_2(7) = 2.8, p = .01$ (one-tailed).

We have claimed that the results of Experiment 1 provide evidence that objects semantically related with a spoken target are activated. However, one might argue that the fixations on semantically related objects were a result of partial activation at the level of word form because of co-occurrence and were not a result of semantic activation. For example, because *lock* and *key* so frequently co-occur in speech, it is possible that there exists a single representation at the level of word form for the term *lock-and-key*. This issue has been raised frequently in the semantic priming literature (e.g., Fischler, 1977; Shelton & Martin, 1992) because of the concern that semantic priming effects could be due to associative factors such as lexical co-occurrence rather than semantic relationships. However, a number of studies have shown that word pairs that are not associated according to free word association tasks (in which people tend to give lexically co-occurring responses) still prime each other if they share enough semantic features (e.g., Perea & Gotor, 1997; McRae & Boisvert, 1998). In our data, when we examined separately the eight object pairs that were not associated according to the University of South Florida free association norms (Nelson, McEvoy, & Schreiber, 1998), the semantic effect remained (Figure 3, right panel). Averaged over the entire trial, this result was significant by both participants, $t_1(29) = 4.2, p < .01$, and items, $t_2(7) = 3.6, p < .01$ (one-tailed).

Thus, with several alternative explanations ruled out, the results of Experiment 1 show that eye movements do reflect the activation of words semantically related to the target. Therefore, despite their lack of phonological overlap with the target and despite the lack of any task-related motivation to attend to them, semantically related words become active enough for their pictorial representations to draw visual attention. This shift of attention is dependent on utterance of the target word and, indeed, is closely time locked to the information available in the speech stream. It is noteworthy that words semantically related to the target start to draw fixations 200–300 ms after target onset, indicating that the semantically related object may have been preferentially fixated on before the target word could be distinguished from words with the same onset. If true, this suggests that eye movements might also be

⁵ We reasoned that this method would be more appropriate than asking participants to judge the visual similarity between the particular target-competitor picture pairs we used because even if the pictures that represented the target and the competitor were highly visually dissimilar (and in fact we selected them to be dissimilar), if the competitor picture was visually similar to the participant's *mental representation* of the target, then fixations on the competitor could be due to its being mistaken for the target. Note that we were less concerned about the converse (i.e., whether the picture of the target was similar to the participant's mental image of the competitor) because mistaking the target for the competitor would reduce the effect of interest.

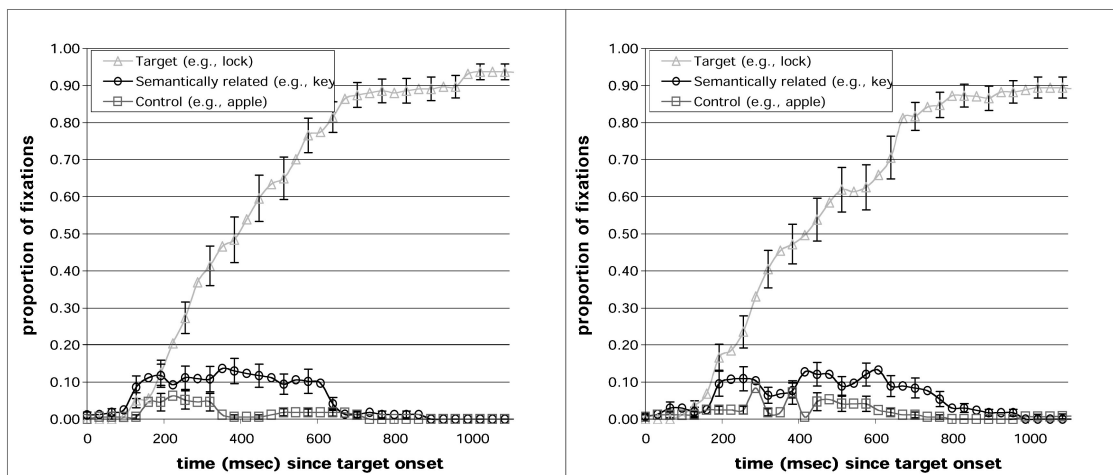


Figure 3. Left panel: The eight object pairs in Experiment 1 with equivalent average shape similarity ratings for the related object versus the target and its control versus the target. Right panel: The eight object pairs in Experiment 1 that are unassociated (forward and backward) according to published free association norms (D. L. Nelson et al., 1998).

drawn to words semantically related to unintended candidates—specifically, to words semantically related to onset competitors of the target. We specifically explore this possibility in Experiment 2.

The results of Experiment 1 compellingly show that increased fixations on the related pictures are triggered by information in the speech stream and do not occur in the absence of the utterance of a semantically related word. However, the data do not preclude the possibility that information gleaned from the pictures prior to the utterance of the target word interacts with the processing of the speech stimulus in such a way as to facilitate mapping of the word onto the referent. For example, it is possible that the picture of the target interacted with the acoustic input as it unfolded, resulting in the target’s activation increasing at a faster rate than it would have as a result of acoustic input alone. This would presumably reduce the amount of acoustic information needed for the target to become active enough to activate the related object.

Examining whether post-word onset fixations on the semantically related object are affected by having previously fixated on the target (pre- or post-word onset) may provide some clues as to whether such an interaction is responsible for the semantic relatedness effect. If it is, post-word onset fixations in trials *with* a previous look to the target should include more fixations on semantically related objects than post-word fixations in trials *without* a previous look to the target. However, we found that in both instances participants looked at the related object in the same proportion of trials. This suggests that the semantic relatedness effect is not an artifact of inflated target activation resulting from preactivation of its semantic representation.

In Experiment 2, we test this claim more explicitly by investigating whether a lexical item that is only temporarily consistent with the speech signal and that is unpictured can become sufficiently activated to result in fixations on a semantically related object. Recall that in Experiment 1 fixations on the semantically related picture began to increase 200–300 ms after word onset, which is well before the target could be distinguished from its onset competitors. This suggests that semantic information is

available soon after word onset. Furthermore, it should be available not only for the target and related words, but also for onset competitors and their related words. Experiment 2 explores whether eye movements are drawn to an object semantically related to an uttered word’s onset competitor—even though the competitor itself is absent from the display.

The design of Experiment 2 allows us to assess the impact of prior presentation of pictorial stimuli on eye movements in response to spoken stimuli and also to investigate whether the semantic information that listeners will activate is limited to information about objects present in the display. The logic of Experiment 2 is similar to that used by Dahan et al. (2001a) who investigated the effects of nonpictured onset competitors on eye movements. As described above, they found that cross-splicing a stimulus word such that its onset came from a real-word competitor created interference, delaying fixations on the target when compared with a stimulus word whose onset was cross-spliced from a nonword. Although the Dahan et al. study demonstrated that listeners do not limit the phonological representations that they consider to those candidates in the immediate visual environment, in light of the robust evidence (described above) that higher level processes do constrain referential interpretation in the visual world paradigm, the question remains open whether semantic information about objects that are not present in the immediate visual environment become active. In Experiment 2, we investigate whether the semantic representation of an unpictured onset competitor can activate a semantically related object strongly enough for it to draw visual attention. Directly comparing the results of our two experiments will allow us to evaluate the extent to which the visual display influences the semantic relatedness effect for eye movements.

Experiment 2

Experiment 1 showed that eye movements to a semantically related object are tightly time locked to the utterance of the target

word. Although the order of fixation analysis in Experiment 1 suggests that previous exposure to the pictures had minimal influence on participants' eye movements, it is still possible that the semantic effect was inflated as a result of participants limiting the candidates that they would consider to those present in the visual display. To address this concern, Experiment 2 investigates whether an unpictured onset competitor of the target word would partially activate a semantically related object. In Experiment 2, we replaced the target objects from Experiment 1 with objects whose names are onset competitors (e.g., we replaced *lock* with *logs*). If words semantically related to onset competitors of the new target words still draw increased eye fixations (despite the fact that the onset competitors themselves are not present in the display) this would constitute strong evidence that the semantic relatedness effect reflects lexical–semantic processing in the general lexicon. More important from a theoretical perspective, it would also demonstrate that the semantic representations of unintended candidates can become active enough to draw visual attention.

Method

The methods used in Experiment 2 were identical to those used in Experiment 1, with the exceptions described below.

Participants. We tested 30 participants (from the Brown University community).

Materials. For 20 of the 24 items used in Experiment 1, the original target was replaced with a phonological onset competitor. For example, the target *lock* was replaced with *logs*. The four remaining items from Experiment 1 were not included in Experiment 2 because a suitable (i.e., picturable and relatively familiar) onset competitor with which to replace the target from Experiment 1 was not available. Three unrelated objects (identified in Appendix B) were replaced to avoid introducing a semantic or visual relationship with the target. No other changes were made to the displays. Target names were one to four syllables long (with an average of two syllables). Average target duration was 569 ms. Two lists, each 40 trials long, were created. As in Experiment 1, each participant was presented with only one list so that no participant saw or heard any item more than once. (See Appendix B for a full list of the Experiment 2 stimuli.) Figure 4 shows a sample display. In the semantic onset competitor condition (10 of 40 trials), one of the pictures in the display was semantically related to an unpictured phonological onset competitor of the target. This object is referred to as the *semantic onset competitor*. The other two objects were semantically and phonologically unrelated to both the target and the semantic onset competitor. The name of one of these unrelated objects (the *competitor's control*) was matched for frequency with the name of the semantic onset competitor. As in Experiment 1, the same displays that were used in the semantic onset competitor condition in one list served (between subjects) as displays in the control condition (10 of 40 trials) on the other list. Thus, every pair of related words that appeared as target and competitor (in the semantic onset competitor condition) on one list appeared as objects unrelated to the target (in the control condition) on the other list. This was accomplished by using one of the objects that served as an "unrelated" object in the semantically related condition as the target in the control condition. (See Appendix B).

In order to ensure that the change in the target would be the only way the displays in Experiment 2 differed from those used in Experiment 1, we did not change the targets in the control condition to frequency match them with targets in the semantic onset competitor condition. The average frequency of targets in the semantic onset competitor condition was slightly lower than the average frequency of targets in the control condition (log frequency = 1.33 vs. 1.58, respectively). The twenty remaining trials were filler trials. There were four practice trials.

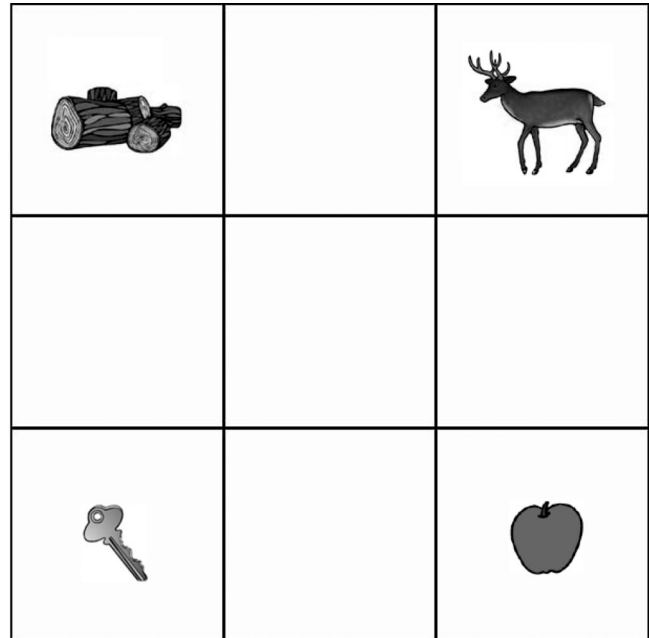


Figure 4. A sample display from Experiment 2. The target object, (*logs*), is an onset competitor of an unpictured object (*lock*), which is semantically related to one of the other objects in the display (the semantic onset competitor, *key*). This same display also appears (between subjects) in the control condition with an unrelated object (*deer*) serving as the target.

Results and Discussion

Data were analyzed as in Experiment 1. One trial (0.2%) was not included in the analysis because the wrong picture was selected. Five percent of trials did not provide any data because there were no eye movements after the onset of the target word (most of these were trials in which the participant was already fixating on the picture of the target at the onset of the target word). Fixation proportions were replaced in the same way as in Experiment 1 when they were more than 2.5 standard deviations away from the mean (3% by participants, 2% by items).

Figure 5 plots the mean proportion of fixations on each picture type over time (from target onset to 1,000 ms after onset) for the semantic onset competitor condition. As in Experiment 1, for the purposes of analyzing the data, a trial was defined as starting 200 ms after target onset and ending at the point at which the probability of fixating on the target asymptoted. In these data the end of the trial occurred at about 900 ms after target onset. When we compared mean fixation proportions averaged over the entire trial, we found that, as predicted, the semantic onset competitor received more fixations than the competitor's control. This difference was significant in the participants analysis, $t_1(29) = 1.9, p = .03$, and in the items analysis, $t_2(19) = 1.7, p = .05$ (one-tailed). The semantic onset competitor effect was also measured by comparing fixations on the semantic onset competitor when it appeared together with the target, to fixations on the same object in the control condition. As expected, the related picture in the semantic onset competitor condition received more fixations than the same picture in the control condition, where it was unrelated to the target, $t_1(29) = 4.5, p < .01$, and $t_2(19) = 5.3, p < .01$ (one-tailed).

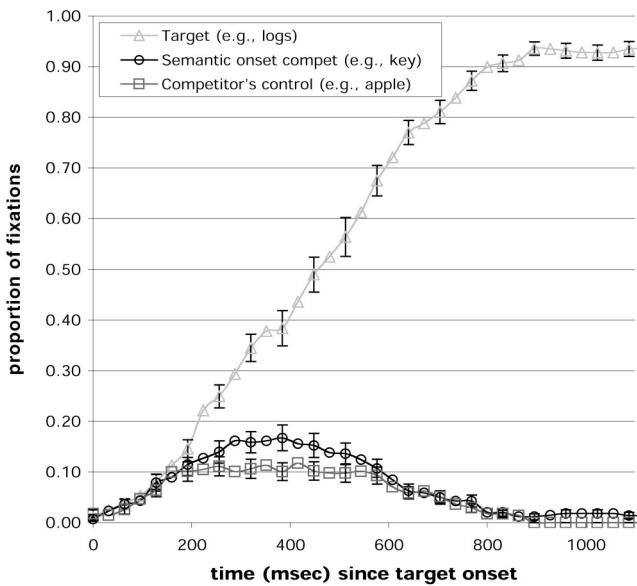


Figure 5. Experiment 2: Proportion of fixations over time on the target, the semantic onset competitor, and the semantic onset competitor's control. Standard error bars are shown for every other data point.

To obtain information about the time course of participants' fixations in the semantic onset competitor condition, we divided the trial into seven 100-ms windows (from 200–900 ms after target onset) and conducted separate planned comparisons on each window. All differences reported below were reliable at $p < .05$ (one-tailed) for both participants and items unless otherwise specified. To establish when the target started to become active, we compared fixations on the target with fixations on the unrelated pictures. In the window from 200 to 300 ms after target onset, there were more fixations on the target picture than on unrelated pictures. Fixations on the target picture remained above those on unrelated pictures until the end of the trial. To determine when the semantic onset competitor became active, we compared fixations on the semantic onset competitor with fixations on the competitor's control picture in the same display. There was no difference between fixations on the competitor and the control picture before 300 ms after target onset. In the windows from 300 to 400 ms and from 400 to 500 ms after target onset there were significantly more fixations on the competitor than on its control ($p < .05$ for participants and items), with one exception: In the items analysis of the 400–500 ms window, the difference between the competitor and the control did not quite reach significance ($p = .07$). The competitor was also fixated on more than its control in the window from 500–600 ms after target onset but this difference only approached statistical significance by participants ($p = .09$) and items ($p = .11$).

The results of Experiment 2 show that words semantically related to phonological onset competitors of an uttered word become active enough to draw visual attention. These findings provide support for the hypothesis that not only the forms, but also the meanings of a word's onset competitors become partially active during spoken word recognition. The results therefore show important convergence with the cross-modal priming literature,

using an experimental paradigm that does not suffer from the same limitations. The results of Experiment 2 also address some of the concerns regarding the validity of the visual world paradigm as a method for studying lexical processing. Of importance, participants preferentially fixated on the semantic onset competitor despite the fact that the onset competitor itself was not present in the display, and none of the pictures in the display were semantically related to any others. Thus the data demonstrate that the activation of semantically related objects can be observed in the visual world paradigm and yet is not driven by the visual domain itself. In addition, they show that eye movements in a task such as this one can be influenced by the activation of representations that do not have corresponding referents in the display. These results are reassuring because they suggest that eye movements can be used to study patterns of lexical and semantic activation that generalize beyond the specific displays used in the experiments.

It is interesting to note that the 200–300 ms duration of the semantic onset competitor effect in Experiment 2 is much shorter than the 700 ms duration of the semantic relatedness effect in Experiment 1. In fact, the short duration of the semantic onset competitor effect suggests that the activation (and deactivation) of the semantic onset competitor is directly linked to the acoustic input. If true, the semantic effect in Experiment 1 should not only last longer than the semantic onset effect in Experiment 2, but the point at which the difference between the two experiments emerges should correspond to the point at which it becomes clear that the target in Experiment 2 is not the target in Experiment 1.

We conducted a gating posttest to measure the isolation point of each target word used in Experiment 2. Word order was randomized for each participant. Twelve participants (who had not participated in Experiments 1 or 2) listened to successively longer segments of the targets from Experiment 2. The initial gate was 120 ms, and each subsequent gate was 30 ms longer until the entire word had been presented. After each gate, participants typed the word they thought they were hearing. The mean isolation point (i.e., how much of the word needed to be heard before it could be identified without any subsequent change in response) for the targets in Experiment 2 was 329 ms after word onset (with a standard deviation of 84 ms). Thus we predicted that although the overall semantic effect in Experiment 1 would be larger (because it would last longer) than the semantic onset effect in Experiment 2, the effects in the two experiments should not differ until about 500–600 ms after target onset (i.e., until after the isolation point plus the approximately 200 ms it takes to launch a saccade).

To test our prediction, for each experiment we computed the difference between the related object and its control, averaged across the entire trial (200–900 ms after target onset). We found that, as predicted, the overall semantic effect in Experiment 1 was larger than the overall semantic onset effect in Experiment 2 for participants $t_1(58) = 2.5, p < .01$, and for items $t_2(42) = 2.6, p < .01$ (both one-tailed). To explore whether this difference emerged at the point at which it became clear that the acoustic input of Experiment 2 was not the target word in Experiment 1, we then divided the trial into 100 ms bins and compared the two experiments in each of those bins. The results of these comparisons are reported in Table 1. Consistent with our prediction, the two experiments did not differ significantly until after 500–600 ms after target onset (though small but not statistically significant differ-

Table 1

Comparison of Semantic Relatedness Effect (Experiment 1) with Semantic Onset Competitor Effect (Experiment 2)

<i>t</i> test results	Bin in ms						
	200–300	300–400	400–500	500–600	600–700	700–800	800–900
$t_s(58)$	0.80	0.30	1.00	1.30	2.50	2.70	2.80
$t_i(42)$	1.00	0.50	1.30	1.20	3.00	3.00	2.50
p_s	0.22	0.39	0.15	0.10	0.007	0.005	0.004
p_i	0.16	0.31	0.10	0.12	0.003	0.003	0.009

Note. All p values are one-tailed.

ences emerged in the bins 400–500 and 500–600 ms after target onset). At 600–700 ms, however, the difference between the two experiments was highly significant, and this difference persisted throughout the remainder of the trial.

This shorter duration of the semantic onset competitor effect is consistent with results from two early cross-modal semantic priming studies (Marslen-Wilson, 1987; Zwitserlood, 1989). In these studies, it was found that if the (written) target was presented early during the presentation of the prime (before the prime's uniqueness point), lexical decisions were speeded to targets related to possible continuations of the prime. However, if the target was presented after the point at which the prime became unambiguous, responses were only speeded if the target was related to the word actually uttered. The onset competitor effect's short duration also supports the account given by some investigators of embedded word activation for why they did not obtain priming for words semantically related to words embedded in the onsets of real words (Isel & Bacri, 1999; Vroomen & de Gelder, 1997). These studies presented visual targets related to onset embedded words at the offset of the embedding words (e.g., they might have presented *lock* as a target at the offset of the prime *keyboard*), and attributed the absence of a priming effect to the "deactivation" of onset embedded words, rather than the absence of activation. Consistent with this account, if in Experiment 2 we had measured only eye movements launched after the target's offset (which would begin to appear at 769 ms, which is 200 ms after the average target offset), we would have found no evidence of the semantic onset competitor's activation.

Thus, Experiment 2 shows that in the visual world paradigm, lexical-semantic processing is not limited to the set of candidates present in the display. Rather, processing occurs in the context of the general lexicon. Together, the results of Experiments 1 and 2 demonstrate that words semantically related not only to the target, but also to its onset competitors become active enough that their pictorial representations draw fixations, despite the absence of any task demands pertaining to these semantically related objects. It is striking that during the portion of the speech stream that was consistent with the onset competitor, there were no reliable differences between the relatedness effects in Experiments 1 and 2. This suggests that lexical activation in response to acoustic input is minimally impacted by the picture displays. In addition, the time course of the effect in Experiment 2 may suggest why, despite the fact that the semantic representations of unintended candidates become active enough to draw visual attention, people are not conscious of diverting attention to them: Their activation is extremely short-lasting.

General Discussion

Summary

The current work shows that eye movements reflect more than the degree of phonological match between the acoustic input and the phonological forms of potential referents in the display. They also reflect the activation of semantic information about candidates that are considered as a spoken word is processed. Experiment 1 shows that visual attention is drawn to objects semantically related to a spoken target word. Experiment 2 shows that objects semantically related to onset competitors of a target word also draw visual attention, at least when these competitors have substantial overlap with the onset of the target. Thus, hearing */la/* not only activates the phonological competitors *lock*, *logs*, and so forth; it also partially activates semantic associates of these words, including *key* and *wood*. It is quite striking that even in an explicitly referential task such as the one used in Experiments 1 and 2, activation of lexical items is not limited to potential referents. This is interesting in light of the numerous studies that have shown rapid effects of linguistic structure and pragmatic information on reference resolution. It suggests that these sources of information do not determine lexical activation levels, rather they interact with a lexical system in which there is considerable bottom-up priority. This bottom-up priority leads to widespread activation, even of lexical items that are unlikely referents. Such rampant activation would seem to threaten chaos, yet the subjective experience of word recognition is not at all chaotic. This suggests that the mechanisms involved maintain a careful balance between activation and deactivation: Widespread activation must be counteracted with swift deactivation.

The results of Experiment 1 converge with the data reported by Huettig and Altmann (2005), who showed that when passively listening to a sentence that refers to an object (e.g., *piano*), people are more likely to fixate on an object from the same conceptual category as the object being referred to (e.g., *trumpet*) than on a picture of an unrelated object. However, unlike in the current work, participants in Huettig and Altmann's study were passively viewing the scene; they were not attempting to pick out the named object, and in fact, displays did not always contain pictures of the target words. The task, therefore, did not carry the explicit expectation of a referential link between the spoken words and pictured referents, unlike Experiments 1 and 2 of the current study. Nevertheless the time course of the effect that they obtained corresponds perfectly to the time course of the effect obtained in Experiment 1, suggesting that the same processes are being reflected in the two studies. Since Huettig and Altmann's study

included only pairs that were not associated according to the University of South Florida free association norms (in which people tend to give lexically co-occurring responses), conceptual rather than lexical relatedness must have been responsible for eye movements to semantically related objects. This result is consistent with our finding (from the subset of unassociated items in Experiment 1) that lexical association is not necessary for semantically related objects to draw visual attention. It should be noted, however, that in Experiment 1 the related stimuli were not limited to the same conceptual category as the target. Also included were pairs related by virtue of having a similar function (e.g., *tape–glue*, *candle–lightbulb*), pairs in which the two objects might interact in a single event (e.g., *lock–key*, *hammer–nail*, *pie–ice cream*) and pairs exhibiting other semantic relations (e.g., *grapes–wine*). This suggests, therefore, that objects standing in a variety of semantic relationships to the target can draw visual attention.

The evident sensitivity of eye movements to the activation of semantic information, coupled with the time course information that they provide, appear to make this paradigm particularly well-suited for investigating the fine details of lexical and semantic activation during spoken word recognition. For example, although the experiments reported in this article made no attempt to control the way in which semantically related objects were related to the target or its phonological competitors (reflecting instead a variety of different relations), by explicitly controlling these semantic relationships one could potentially tap into the organization of semantic memory. In particular, by varying the semantic relationship between the target and the semantically related object, the paradigm introduced here could be used to obtain information about whether semantic memory is organized such that words that share particular semantic features partially activate each other. Along these lines, Myung, Blumstein, & Sedivy (in press) have recently demonstrated that visual attention is drawn by related objects that share the manner in which they are manipulated with target objects, even when they share no other robust semantic relation (e.g., *typewriter–piano*). Similarly, other researchers have shown that distractor objects that are similar in shape to a target object draw visual attention (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2004). Moreover, because it has been proposed that we do not access all types of semantic information about a word at the same time but rather that different kinds of semantic information (e.g., perceptual and functional) become active on different time courses (Schreuder, Flores d'Arcais, & Glazenbourg, 1984), the time course information that the eye movement paradigm provides may make it particularly valuable for investigating semantic activation.

Because of its simplicity and the time course information that it provides, the eye tracking paradigm is also promising as a tool for investigating the time course of lexical and semantic activation in brain damaged populations. For instance, theories attempting to account for the lexical processing deficits in Broca's and Wernicke's aphasia make predictions that relate to the time course of lexical activation (e.g., Hagoort, Brown, & Swaab, 1996; Prather, Zurif, Love, & Brownell, 1997; Milberg, Blumstein, & Dworetzky, 1988). Although numerous studies have explored the time course of lexical activation in aphasic patients by varying the interstimulus interval in a semantic priming paradigm (Hagoort, 1993; Prather, Zurif, Stern, & Rosen, 1992; Prather, Zurif, Love, & Brownell, 1997), information about the time course of lexical

activation in Broca's and Wernicke's aphasia remains spotty because in the semantic priming paradigm activation is probed at discrete points. Because eye tracking can provide a more direct and continuous measure of lexical activation, using this paradigm could provide insight into the lexical processing deficits of these patients. Studies of eye movements in patients with lexical processing deficits (Yee, Blumstein, & Sedivy, 2004) may not only shed light on their disorders but could also provide clues about how a properly functioning system maintains the delicate balance between activation and deactivation that was demonstrated in Experiment 2.

Conclusion

The present findings indicate that during spoken word recognition, words semantically related to the uttered word and words semantically related to the uttered word's onset competitors become active enough to draw visual attention. People's visual attention is diverted despite the fact that they are rarely aware of activating the meanings of candidates that turn out to be incorrect, and even though attending to these semantically related objects does not help (and in fact may hinder) task performance. These findings provide support for models of spoken word recognition that claim that when a word's form is even partially activated, its meaning is also automatically activated. How is it possible to account for the fact that the meanings of unintended candidates are activated in the absence of conscious awareness? One important clue may come from the short duration (~200 ms) of the semantic onset competitor effect—it could be that this activation is too brief to reach conscious awareness. The short duration of this effect also suggests that (a) the semantic activation of unintended candidates quickly decays when the input is no longer consistent with their forms, and/or (b) more active candidates inhibit less active candidates.

These findings also provide important validation of the visual world paradigm: Semantic relatedness effects occur regardless of whether the lexical item that is the source of the semantic activation is pictured in the display. Thus, this method can be used to productively measure the time course over which a word's meaning becomes active (as well as its form). At the same time this method has the temporal sensitivity and ecological validity to provide detailed time course information about lexical processes in various populations.

References

- Alloppenna, P. D., Magnuson, J., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence of continuous mapping models. *Journal of Memory and Language*, *38*, 419–439.
- Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action*. New York: Psychology Press.
- Bajo, M.-T. (1988). Semantic facilitation with pictures and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 579–589.
- Bates, Timothy, C., & Oliveiro, L. (2003). PsyScript: A Macintosh application for scripting experiments. *Behavior Research Methods, Instruments, & Computers*, *35*, 565–576.

- Carr, T. H., McCauley, C., Sperber, R. D., & Parmelee, C. M. (1982). Words, pictures, and priming: On semantic activation conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 757–777.
- Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2001). Circumscribing referential domains in real-time language comprehension. *Journal of Memory and Language*, *47*, 30–49.
- Connine, C. M., Titone, D., Dellman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*, 463–480.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001a). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534.
- Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001b). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*, 317–367.
- Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*, 453–459.
- Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition*, *73*, B1–B15.
- Durso, F. T., & Johnson, M. (1979). Facilitation in naming and categorizing repeated pictures and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *5*, 449–459.
- Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & Cognition*, *5*, 699–716.
- Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*, 631–656.
- Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development, *IEEE ICASSP*, 517–520.
- Hagoort, P. (1993). Impairments of lexical-semantic processing in aphasia: Evidence from the processing of lexical ambiguities. *Brain and Language*, *45*, 189–232.
- Hagoort, P., Brown, C. M., & Swaab, T. Y. (1996). Lexical-semantic event-related potential effects in patients with left hemisphere lesions and aphasia and patients with right hemisphere lesions without aphasia. *Brain*, *119*, 627–649.
- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43–61.
- Huetting, F., & Altmann, G. T. M. (2004). The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In M. Carreiras, & C. Clifton (Eds.), *The online study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 187–207). New York: Psychology Press.
- Huetting, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*, B23–B32.
- Isel, F., & Bacri, N. (1999). Spoken-word recognition: The access to embedded words. *Brain and Language*, *86*, 61–67.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words, the neighborhood activation model. *Ear and Hearing*, *19*, 1–36.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition*, *25*, 71–102.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.
- McNellis, M. G., & Blumstein, S. E. (2001). Self-organizing dynamics of lexical access in normals and aphasics. *Journal of Cognitive Neuroscience*, *13*, 151–170.
- McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 558–572.
- Milberg, W., Blumstein, S., & Dworetzky. (1988). Phonological processing and lexical access in aphasia. *Brain and Language*, *34*, 279–293.
- Mitchell, M. P., Santorini, B., & Marcinkiewicz, M. A. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* *19.2*, 313–330.
- Moss, H. E., McCormick, S. F., & Tyler, L. K. (1997). The time course of activation of semantic information during spoken word recognition. *Language and Cognitive Processes*, *12*, 695–731.
- Myung, J., Blumstein, S. E., & Sedivy, J. C. (in press). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition*.
- Nadig, A., & Sedivy, J. (2002). Evidence of perspective-taking constraints in children's online reference resolution. *Psychological Science*, *13*, 329–336.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.
- Perea, M., & Gotor, A. (1997). Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition*, *62*, 223–240.
- Prather, P., Zurif, E. B., Stern, C., & Rosen, J. T. (1992). Slowed lexical access in nonfluent aphasia: A case study. *Brain and Language*, *43*, 336–348.
- Prather, P. A., Zurif, E., Love, T., & Brownell, H. (1997). Speed of lexical activation in nonfluent Broca's aphasia and fluent Wernicke's aphasia. *Brain and Language*, *59*, 391–411.
- Rossion, B. & Pourtois, G. (2001, May). Revisiting Snodgrass and Vanderwart's object database: Color and texture improve object recognition. Paper presented at the *VisionSciences* meeting, Sarasota, Florida.
- Scarborough, D. L., Gerard, L., & Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory & Cognition*, *7*, 3–12.
- Schreuder, R., Flores d'Arcais, G. B., & Glazenborg, G. (1984). Effects of perceptual and conceptual similarity in semantic priming. *Psychological Research*, *45*, 339–354.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109–147.
- Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1191–1210.
- Smith, M. C., Meiran, N., & Besner, D. (2000). On the interaction between linguistic and pictorial systems in the absence of semantic mediation: Evidence from a priming paradigm. *Memory & Cognition*, *28*, 204–213.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–216.

Spivey, M., Tanenhaus, M., Eberhard, K., & Sedivy, J. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45, 447–481.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995, June 16). Integration of visual and linguistic processing in spoken language comprehension. *Science*, 268, 1632–1634.

Tweedy, J. R., Lapinski, R. H., & Schaneveldt, R. W. (1977). Semantic-context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition*, 5, 84–89.

Vanderwart, M. (1984). Priming by pictures in lexical decision. *Journal of Verbal Learning and Verbal Behavior*, 23, 67–83.

Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 710–720.

Yee, E., Blumstein, S. E., & Sedivy, J. (2004). The time course of lexical activation in Broca's and Wernicke's aphasia: Evidence from eye movements. *Brain and Language*, 91, 62–63.

Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, 25–64.

Appendix A

Experiment 1

Stimuli Used in Experiment 1

List	Target (Unrelated)	Semantically related (Control for related in other list)	Target's control (Target)	Related's control (Unrelated)
A	lock	key	deer	apple
A	bat	racket	stove	pear
A	battery	plug	ghost	map
A	car	bike	money	scale
A	grapes	wine	cane	bell
A	mitten	glove	seahorse	bucket
A	muffin	donut	camel	lighthouse
A	piano	trumpet	belt	lamp
A	pie	ice cream	chest	dice
A	sock	shoe	peach	racket
A	tape	glue	penny	celery
A	teepee	igloo	hanger	paperclip
B	candle	lightbulb	suitcase	headphones
B	cat	mouse	bed	pump
B	hammer	nail	monkey	couch
B	matches	lighter	sandwich	bonnet
B	pants	shirt	cake	tire
B	robe	slippers	olives	canoe
B	saw	axe	pig	vest
B	scissors	knife	ashtray	coat
B	telescope	binoculars	squirrel	pumpkin
B	tie	jacket	paint	cow
B	wallet	purse	lobster	drum
B	window	door	football	gun

Note. Rows depict displays. Each display was presented to participants receiving Lists A and B. Open column headers indicate the assignment of objects to conditions for the list indicated in column 1, whereas parenthetical headers indicate the assignment of objects to conditions for the other list. For example, for participants receiving List A, *lock* was the target, *key* was the semantically related object, *deer* was the target's control, and *apple* was the related object's control. Participants receiving List B saw the same display, but *deer* was the target, *key* was a control for the related object in list A, and *apple* and *lock* were unrelated objects.

Appendix B

Experiment 2

Stimuli Used in Experiment 2

List	Target (Unrelated)	Competitor (Control for competitor in other list)	Unrelated (Target)	Competitor's control (Unrelated)
A	logs (lock)	key	deer	apple
A	bat (battery)	plug	money ^a	map
A	cards (car)	bike	ghost ^a	scale
A	grapefruit (grapes)	wine	cane	bell
A	muffler (muffin)	donut	camel	lighthouse
A	peanut (piano)	trumpet	belt	lamp
A	soccer ball (socks)	shoe	peach	beer
A	table (tape)	glue	penny	celery
A	teapot (teepee)	igloo	hanger	paperclip
A	walrus (wallet)	purse	lobster	drum
B	battery (bat)	racket	suitcase	pear
B	candy (candle)	lightbulb	headphones	paperclip
B	caterpillar (cat)	mouse	bed	pump
B	hammock (hammer)	nail	monkey	chocolate ^b
B	mattress (matches)	lighter	sandwich	bonnet
B	panther (pants)	shirt	cake	tire
B	rope (robe)	slippers	olives	canoe
B	sock (saw)	axe	pig	vest
B	telephone (telescope)	binoculars	squirrel	pumpkin
B	windmill (window)	door	football	gun

Note. Rows depict displays. Each display was presented to participants receiving Lists A and B. Open column headers indicate the assignment of objects to conditions for the list indicated in column 1, whereas parenthetical headers indicate the assignment of objects to conditions for the other list. For example, for participants receiving List A, *logs* was the target, *key* was the semantic onset competitor, and *apple* was the competitor's control. Participants receiving List B saw the same display, but *deer* was the target, *key* was a control for the semantic onset competitor in list A, and *apple* and *logs* were unrelated objects. Unpictured onset competitors are indicated in parentheses.

^a *Ghost* and *money* were exchanged because they would have been related to the (new) target in their original positions.

^b *Couch* was replaced with *chocolate* because *couch* would have been related to the (new) target.

Received January 23, 2005
 Revision received June 23, 2005
 Accepted July 22, 2005 ■