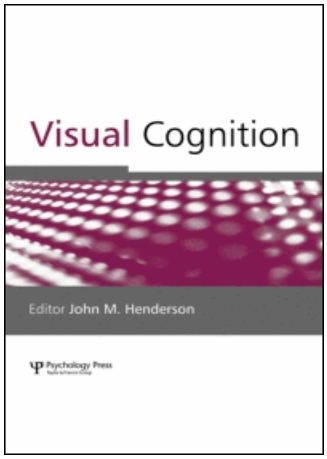

Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683696

Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness

Falk Huettig [a]; Gerry T. M. Altmann [b]
[a] Ghent University, Belgium
[b] University of York, UK

Online Publication Date: 01 November 2007
To cite this Article: Huettig, Falk and Altmann, Gerry T. M. (2007) 'Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness', Visual Cognition, 15:8, 985 - 1018
To link to this article: DOI: 10.1080/13506280601130875
URL: http://dx.doi.org/10.1080/13506280601130875

VISUAL COGNITION, 2007, 15 (8), 985–1018

Psychology Press
Taylor & Francis Group

Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness

Falk Huettig
Ghent University, Belgium

Gerry T. M. Altmann
University of York, UK

Visual attention can be directed immediately, as a spoken word unfolds, towards conceptually related but nonassociated objects, even if they mismatch on other dimensions that would normally determine which objects in the scene were appropriate referents for the unfolding word (Huettig & Altmann, 2005). Here we demonstrate that the mapping between language and concurrent visual objects can also be mediated by visual-shape relations. On hearing "*snake*", participants directed overt attention immediately, within a visual display depicting four objects, to a picture of an electric cable, although participants had viewed the visual display with four objects for approximately 5 s before hearing the target word—sufficient time to recognize the objects for what they were. The time spent fixating the cable correlated significantly with ratings of the visual similarity between snakes in general and this particular cable. Importantly, with sentences contextually biased towards the concept snake, participants looked at the snake well before the onset of "*snake*", but they did not look at the visually similar cable until hearing "*snake*". Finally, we demonstrate that such activation can, under certain circumstances (e.g., during the processing of dominant meanings of homonyms), constrain the direction of visual attention even when it is clearly contextually inappropriate. We conclude that language-mediated attention can be

Please address all correspondence to Falk Huettig, now at the Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands.
E-mail: Falk.Huettig@mpi.nl

FH was supported by a University of York doctoral studentship. The eyetracking facilities and software/analysis routines were supported by Medical Research Council grants G9628472N and G0000224 awarded to GA. Additional support was provided to FH by a grant from FWO, the Fund for Research Flanders (G.0181.05). Aspects of these data were presented at the AMLaP (2002) conference in Tenerife and the 26th annual meeting of the Cognitive Science Society (2004). A partial report of these data appears in Huettig and Altmann (2004). The authors would like to thank Philip Quinlan, Gareth Gaskell, and Graham Hitch for their valuable contributions during the course of this research. We also thank Delphine Dahan, John Henderson, and two anonymous reviewers for helpful comments on an earlier version of this paper.

http://www.psypress.com/viscog DOI: 10.1080/13506280601130875

guided by a visual match between spoken words and visual objects, but that such a match is based on lexical input and may not be modulated by contextual appropriateness.

Casual conversation is typically conceived of as an effortless activity yet, even on superficial analysis, securing a mapping between the surface structure of a given word and its intended meaning can be far from transparent. Recently, the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) has begun to illuminate some very basic questions concerning the online interpretation of speech and its integration with visual context (see Henderson & Ferreira, 2004, for recent review).

Cooper (1974) showed in an early visual world study that participants tended to spontaneously fixate the visual referents of words concurrently heard. For instance, they were more likely to fixate the picture of a snake when hearing "*snake*"[1] or part of "*snake*" than pictures of referents of unrelated control words (see also Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001). Moreover, participants were more likely to fixate pictures showing a snake, a zebra, or a lion when hearing the semantically related word "*Africa*" than they were to fixate referents of semantically unrelated control words. In this respect, the pattern of eye movements reflected the online activation of word semantics and its integration with concurrent visual objects (see also Yee & Sedivy, 2006, for a similar result). We (Huettig & Altmann, 2005) recently further pursued this finding by investigating whether semantic properties of individual lexical items could direct eye movements towards objects in the visual field in the absence of any associative relationships between the words heard and the concurrent visual objects. Our participants were shown a visual display containing four pictures of common objects. During the course of a trial, a spoken sentence was presented to the participant and the participant's eye movements were tracked as the sentence unfolded. We found that participants directed overt attention immediately towards a picture of an object such as a trumpet when a semantically related but nonassociated target word (e.g., "*piano*") was heard. Three different measures of semantic relatedness (McRae feature norms, Cree & McRae, 2003; LSA, Landauer & Dumais, 1997; contextual similarity, McDonald, 2000) each separately correlated well with fixation behaviour (Huettig & Altmann, 2005; Huettig, Quinlan, McDonald, & Altmann, 2006). These data suggest that language-mediated eye movements to objects in the concurrent visual

[1] Double quotes and italic font are used to denote spoken language materials.

environment are driven by semantic similarity in addition to associative knowledge.

Given the observed *conceptual* synergies, it is important to explore *perceptual* synergies in a similar fashion. Little attention has focused on examining the interaction of spoken language with directed attention and the *visual* properties of the presented objects. In this regard, Cooper (1974) also found that participants tended to fixate a picture of a snake when hearing the word "*wormed*" (in the context "*just as I had wormed my way on my stomach*"). This finding (although not discussed by Cooper) suggests that there may also be a strong link between lexical processing and the visual properties of an object such as an object's shape (although it cannot be ruled out that in Cooper's experiment participants mistook the snake for a worm and therefore directed their attention to the picture of the snake when hearing "*wormed*").

The question of whether perceptual information (such as an object's visual shape) becomes automatically available on hearing a spoken word such as "*coin*" has attracted some attention in language comprehension research, although the data have at times been contradictory. Schreuder and colleagues (e.g., Flores d'Arcais, Schreuder, & Glazenborg, 1985; Schreuder, Flores d'Arcais, & Glazenborg, 1984) obtained significantly facilitated target naming times for perceptually related word pairs (*button-coin*), and proposed a model of lexical activation in which perceptual representations are activated very rapidly during spoken word recognition. However, Pecher, Zeelenberg, and Raaijmakers (1998) found perceptual priming only if participants were first given practice in categorizing primes and targets in a perceptual categorization task. Moss, McCormick, and Tyler (1997) reported a significant priming effect for perceptual targets (e.g., pairs with a similar visual-shape such as *hook* and *curve*) using a lexical decision task. Kellenbach, Wijers, and Mulder (2000) obtained robust perceptual priming as indexed by the ERP N400 component but, contrary to Moss et al., observed no effect for the same materials from the ERP study when used in a lexical decision task.

The lack of consensus in the comprehension literature suggests that the visual world paradigm may be particularly useful for investigating how rapidly, and under what circumstances, visual-shape information becomes available during spoken word processing. This is because of the closely time-locked, fine-grained effects the method has been shown to provide (e.g., Allopenna et al., 1998; Dahan et al., 2001; Salverda, Dahan, & McQueen, 2003).

Recently, Dahan and Tanenhaus (2005) explored how the probability of fixating a rope depicted within a display of four objects changed as participants in a "visual-world" study were instructed to use a mouse to click on a snake depicted in the same visual display. Participants would hear a single word, such as "*snake*", and had to click on the snake. Dahan

and Tanenhaus found that the rope was fixated more often, from between 200 and 300 ms post word onset, than the distractors (e.g., an umbrella and a couch). They concluded that "participants could orient their gaze toward an object's spatial location because its structural representation matches the visual representation of the concept activated by the phonetic input" (p. 457).

At first glance, such a result may seem unsurprising. However, it rules out two possibilities regarding the nature of the process that guides the eyes towards named objects in the visual field. First, and as pointed out by Dahan and Tanenhaus (2005), it rules out the possibility that this guidance is based merely on phonological information associated with the visual objects—that is, it could not be the case that participants are covertly naming objects in the visual field, and then using the match between these names and the unfolding speech stream to guide eye movements towards whichever phonological matches are so obtained (this possibility is also ruled out by our earlier semantic competitor effects; Huettig & Altmann, 2004, 2005). Second, it rules out the possibility, permitted by our earlier competitor effects, that guidance is based only on semantic fit between the concepts activated by the visual objects and the concepts activated by the objects referred to in the speech stream. In principle, knowing that there is a trumpet in the visual field, and hearing "*trumpet*", may cause reorientation towards the trumpet without the need for visual form information to mediate that orientation process; the same conceptual-semantics that causes orientation towards the trumpet when hearing "*piano*" would instead be responsible—visual form would be implicated only in the process that mediates between experiencing the visual image of the object and activating that object's associated conceptual semantics. The idea that guidance might be purely semantic is not implausible given the unreliability of visual form (due to changes in perspective and the fact that, in the dynamically changing world, objects' forms can change on a moment-by-moment basis).

Although compelling, the Dahan and Tanenhaus (2005) study leaves a number of questions unanswered. First, the snake and its visual competitor, the rope, were copresent in the display. But would the same effects be obtained if the rope was the only item that visually resembled a snake in the display? As we shall argue below, this is a critical case that allows us to explore in more detail Dahan and Tanenhaus' conclusion that "finding the referent of a linguistic expression in a circumscribed context is similar to that of natural visual search in which the 'top-down' target representation, activated from the spoken input, is mapped onto the 'bottom-up' scene representation" (p. 457). It is certainly true that in a task in which participants have to click on a named object, an element of visual search may be required (see Vickery, King, & Jiang, 2005, for research within the standard visual search paradigm). However, as Dahan and Tanenhaus

themselves point out, their conclusion may not hold for other visual world tasks. In other versions of the paradigm, we suspect that an alternative view of the relationship between linguistic expressions and scene representation is required: one in which the "top-down" target representation is visually derived (i.e., a picture-derived representation), and it is the "bottom-up" spoken input (i.e., a language-derived representation) onto which this is mapped.

In the visual world paradigm, objects are generally presented in the participant's visual field *before* the referring expressions that might pick out one object or another. Thus, by the time the target linguistic expression (e.g., "*snake*") is encountered, conceptual representations corresponding to the objects in the visual field will already have been activated. These representations would include episodic knowledge regarding the location of the objects in the visual field and their actual form, as well as semantic knowledge about these objects (which may include their "prototypical form"—the visual form derived from the visual scene may be subject to distortions of perspective). Thus, we believe that the picture-derived representation has no less a "top-down" influence than have the language-derived representations due to the spoken input; language-derived conceptual representations are no more mapped onto the visual display than are picture-derived conceptual representations mapped onto the language.

One of the aims of the present research is, within the context of visual competitor effects, to distinguish between an account based on language activating "top-down" representations that are subsequently mapped onto "bottom-up" visual display information (cf. Dahan & Tanenhaus, 2005) and an account based on language representations modulating the activation of (already activated) conceptual visual display-based representations. To do this, we shall ask how visual competition might be mediated by linguistic context. For example, in the context of the sentence fragment "*the man was worried, but then he saw the snake*", how would looks towards the snake (in some appropriate task) differ as compared with the case "*the zookeeper was worried, but then he saw the snake*"? We predict that there will be *more* looks towards the snake even before the word "*snake*" was heard, just because of the relationship between zookeepers and snakes (cf. the semantic "priming" effect reported by Huettig & Altmann, 2005). Perhaps hearing "*zookeeper worried greatly*" activates to some degree the conceptual representation associated with snakes, and this drives the eyes to whatever in the visual scene matches the featural specification associated with snakes (both semantic and physical). But how would this *bias* affect looks towards a rope (or a cable) *in the absence of a snake*? If hearing "*zookeeper worried greatly*" activates the conceptual representation associated with snakes, and if this in turn activates visual representations associated with snakes, we should see increased looks towards the rope in such snake-biasing contexts.

Alternatively, when the snake is present (and not the rope), any increased bias to look towards the snake may not be because "*zookeeper*" activates snake representations (and all that they entail), but because the pre-existing representation of the specific snake depicted in the scene is boosted on hearing the semantically related "*zookeeper*" (cf. Huettig & Altmann, 2004)—in other words, without that pre-existing representation of the specific snake, "*zookeeper*" may not activate the *form* of any specific animal, but may instead activate semantic features associated with animals more generally. If this is the case, and in the absence of a snake in the visual scene, hearing "*zookeeper*" should not engender more looks towards a rope. Only when the word "*snake*" is encountered should the interaction between the scene representation of the rope and the conceptual representation evoked by "*snake*" conspire to drive eye movements towards the rope.

One of the aims of the present research was to investigate precisely such effects, and to explore how the concepts activated by spoken words and visual objects might interact (as evidenced by eye movements) in differing contexts. In Experiment 1, we measured eye movements to (a) the target picture (e.g., a picture of a snake) when participants heard a sentence that was contextually neutral up to the point when the target word was heard (e.g., "*In the beginning, the man watched closely, but then he looked at the snake and realized that it was harmless*"); (b) the target picture (e.g., the snake) when participants heard a sentence that was contextually biased towards the target (e.g., "*In the beginning, the zookeeper worried greatly, but then he looked at the snake and realized that it was harmless*"); and (c) when participants heard the same sentence as in the biasing condition (e.g., "*In the beginning, the zookeeper worried greatly, but then he looked at the snake . . .*") but the target picture (the snake) was replaced by a visual-shape competitor (e.g., a picture of an electric cable). Aside from its theoretical relevance (see above), this last condition has the added advantage relative to the Dahan and Tanenhaus (2005) case that the presence of the cable in the absence of a snake draws attention away from (or rather does not drive attention towards) the physical similarity between cables and snakes. In principle, the biasing context should not "favour" any of the objects depicted in the scene (although this is an empirical issue we return to below). The neutral condition was included in order to establish a baseline against which the efficacy of the biasing contexts 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 was successful), an advantage for the target object should be observed prior to the target word. Our rationale for presenting the visual-shape 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).

Experiment 2 explored further the conditions under which visual form effects occur. The existence of lexically ambiguous words such as "pen" (the writing instrument or the enclosure) enables cases in which the target and competitor are related to alternative meanings of the same phonological word. A context that biases towards one interpretation of the homonym (e.g., towards the enclosure meaning of "pen") might function in much the same way as the zookeeper-snake contexts of Experiment 1. Thus, a sentence fragment such as "*the welder locked up carefully, but then he checked the pen . . .*" might cause increased looks towards a depiction of a pen-enclosure even before the word "*pen*" is encountered. But at "*pen*", two distinct representations become activated—one associated with the enclosure meaning, and the other associated with the writing implement meaning. A variety of studies (e.g., Swinney, 1979; see Simpson, 1994, for review) suggest that the linguistic context will not prevent activation of the unintended meaning. For polarized homonyms such as "pen", where one meaning (writing implement) is more frequent than the other (enclosure), the more frequent (or "dominant") meaning would normally become the most active, although a context biasing the less frequent ("subordinate") meaning may boost the activation of the representation associated with this meaning and bring it up to the level of the more dominant representation (cf. Duffy, Morris, & Rayner, 1988). Thus, in a neutral linguistic context and a visual context depicting both a pen-as-writing-implement and a pen-as-enclosure, we could expect the word "*pen*" to engender more looks to the pen-as-writing-implement than to the pen-as-enclosure, but more looks to the pen-as-enclosure than to unrelated distractors (on the assumption that visual contexts are no different in respect of their influence on lexical ambiguity than linguistic contexts). In a biasing linguistic context (biasing towards the subordinate meaning), we might expect that by the time "*pen*" is heard, there would be more looks towards the pen-as-enclosure, but that at or soon after "*pen*", looks towards the pen-as-writing-implement would rise rapidly. But what if the pen-as-writing-implement were replaced with a visual competitor, i.e., an object that shared its visual form? Would looks towards a sewing needle rise in this same way?

On the one hand, if looks towards the sewing needle did rise on hearing "*pen*", this would suggest that visual attention was driven by the partial featural match between the episodic representation of the needle (including its visual form) and the conceptual representation (including prototypical visual form) associated with one of the meanings of the lexically ambiguous "pen".

On the other hand, the depicted sewing needle is not a visual competitor for the contextually intended pen-as-enclosure (in the way that the cable is a

visual competitor in Experiment 1 for the contextually intended snake)—it is a visual competitor for the contextually *unintended* pen-as-writing-implement. Moreover, the representation of the *intended* pen-as-enclosure may be activated well before the onset of the word "*pen*". Thus, whereas Experiment 1 investigates looks towards an object (the cable) sharing visual form with the visual representation associated with a contextually appropriate concept (activated at the word "*snake*"), Experiment 2 investigates looks towards an object (the needle) sharing visual form with the visual representation associated with a contextually less relevant concept (activated at the word "*pen*"). Thus, if we found that there were no more looks towards the sewing needle than towards the distractors, we would have to conclude that visual competitor effects are modulated by the contextual appropriateness of the target concept (i.e., the concept that conveys prototypical form information shared with the actual form of the visual competitor).

In sum, the primary aim of the present research was to investigate how the concepts activated by spoken words and visual objects interact in differing contexts and how this may be mediated by a match in visual form. Participants saw visual displays containing four spatially distinct objects and heard spoken sentences while their eye movements were recorded. In Experiment 1, the sentences were neutral or were biased such that they contextually biased ("*zookeeper worried greatly*") a certain target word (e.g., "*snake*"). The visual display included the target object (the snake) or an object with a similar visual form (an electric cable). We investigated whether on hearing "*snake*" participants would shift overt attention to an object with a similar visual form (the cable; cf. Dahan & Tanenhaus, 2005). Of particular interest though was whether participants would look at the cable in the snake-biasing context even before hearing "*snake*". In other words, we are asking whether the visual form effect is lexically driven or whether it is modulated by linguistic context. In Experiment 2, we further explored this issue and also used neutral and biasing sentences; the target words, however, were homonyms, and the contextual bias was towards the subordinate meaning of the homonym. The visual display included a depiction of the subordinate meaning and a depiction of the dominant meaning or an object with a similar visual form as the dominant meaning of the homonym. Of particular interest was whether and when participants would look at the visual competitor of the contextually inappropriate dominant meaning. Thus, the present research investigates the conditions under which attention is guided by a visual match between spoken words and concurrent visual objects. We examine whether this visual match is primarily based on lexical input or whether it is modulated by the appropriateness of the linguistic context, and we explore whether anticipatory eye movements can be modulated by a visual match or are based primarily on a semantic match.

EXPERIMENT 1

Method

Materials. On each trial in the experiment, participants were presented with a visual display containing line drawings of four spatially distinct objects together with a spoken sentence. Throughout the experiment the participant's direction of eye gaze was measured. Two sets of 21 visual stimuli were created. In one set (the target set, see Figure 1a), each display contained a target picture of a named item, e.g., the picture of a snake, together with three distractor pictures of objects completely unrelated to any of the spoken words. Importantly, all of these distractors were visually different to the target word's referent. In the second set of stimuli (the competitor set), the target picture was replaced with a picture depicting an object of a similar shape to that of the target's referent (e.g., the picture of a cable, see Figure 1b).

The visual stimuli were selected from commercially available ClipArt packages and presented in greyscale format. The 21 target-competitor pairs were: anchor/arrow, apple/moon, banana/sword, bell/hat, button/coin, candle/tube, cigar/carrot, chimney/rocket, dice/ice cubes, football/planet, globe/orange, horseshoe/magnet, lighthouse/flask, microphone/cone, mirror/frame, pencil/column, plate/wheel, racket/saucepan, scissors/chopsticks, snake/cable, and wheelbarrow/sledge. Naming agreement on all pictures was collected from 46 participants. Responses were coded as intended, unintended, or "no response". Responses were coded as intended only when they exactly matched the intended name (e.g., "bear" was coded as valid but not "Grizzly bear"). "No response" was given in only 0.10% of trials. The intended response was given in 75% of trials. Unintended names were largely due to choosing a near-synonym ("lead" instead of "cable" or "boat" instead of "submarine"). Unintended names due to misidentification occurred on only 1.97% of trials.

There were three experimental conditions: In the neutral condition, target set pictures were each paired with a neutral sentence such that the sentence could not induce any bias to look towards any particular picture in the display, e.g., "*In the beginning, the man watched closely, but then he looked at the snake and realized that it was harmless.*" In the biasing condition, different sentences from the target set were each paired with a sentence that was constructed so as to bias looks towards the designated target picture (e.g., the snake): "*In the beginning, the zookeeper worried greatly, but then he looked at the snake and realized that it was harmless.*" The neutral sentences were included as baseline.

The biasing condition was included in order to investigate the effect of the sentential bias on looks to the target picture (the snake). However, the verb

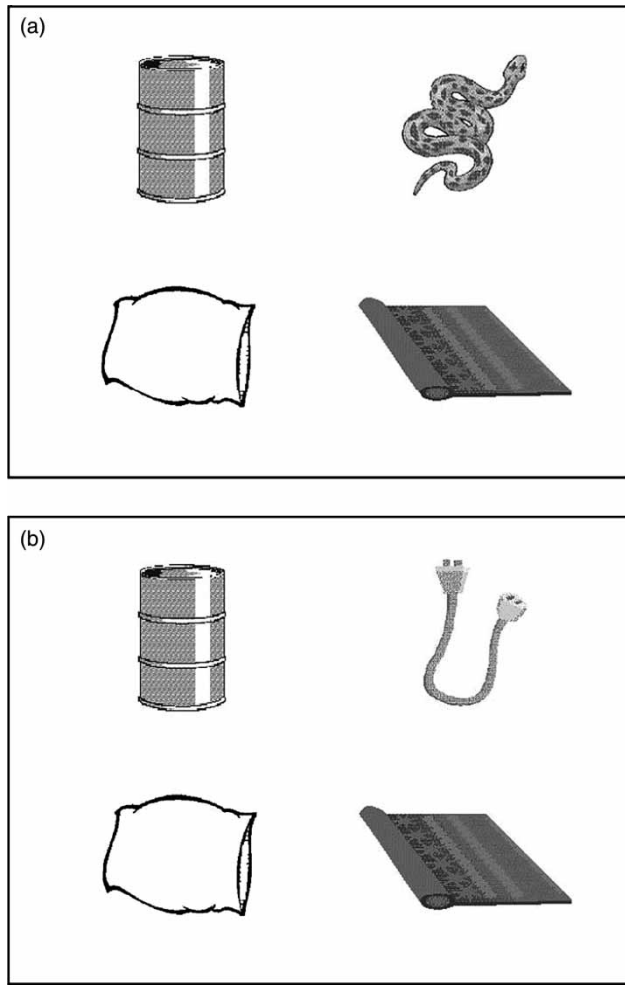

Figure 1. Examples of visual displays used in Experiment 1: (a) In the neutral and biasing conditions, and (b) in the competitor condition. In (a) the target is snake, and in both cases the pictures of the rug, the pillow, and the barrel are distractors; in (b) the cable is the visual-shape competitor.

phrase (e.g., "*looked at*") preceding the target word (e.g., "*snake*") was neutral with respect to the depicted objects. This was to avoid any possible confounding effects that a biasing verb may have had on participants' eye movements.

In each of these first two conditions, the visual display contained the target picture (e.g., a snake). However, in the third and final "shape competitor" condition, the same sentences as in the biasing condition were

paired with the visual competitor set. For these stimuli, the visual target (e.g., the snake) was replaced with a visual competitor (a cable). These competitors were chosen so as to be semantically unrelated to the spoken target word (i.e., "*snake*"). Therefore, the sentence could not induce any bias towards any particular picture. The shape competitor was presented in the biasing context in order to make it even more unlikely that participants would anticipate, prior to the target word, that the shape competitor was the designated special picture.

Preliminary norming study—shape similarity

In a preliminary norming study 12 participants provided relevant ratings. Participants were presented with the written target word (e.g., snake) together with pictures used in the later visual world study, and they were asked to judge how similar the typical physical shape of the target referent was to the physical shape of the referents of the depicted objects. Participants were asked to judge the shape similarity on a scale from 0 to 10 (zero representing "absolutely no similarity in physical shape", 10 representing "identical in physical shape").

The mean rating for the shape competitors was 7.1 ($SD = 1.8$) and 1.4 ($SD = 0.7$) for the distractors. These differences in the shape similarity judgements between the shape competitors and the distractors were highly significant, $F1(1, 11) = 268.89$, $MSE = 0.07$, $p < .01$; $F2(1, 20) = 200.35$, $MSE = 0.17$, $p < .001$. Therefore, the competitor pictures were judged to be significantly more similar in physical shape to the target referents than to those of the distractor pictures.

Visual world study

Participants. 48 participants from the University of York student community took part in this study. All were native speakers of British English and had either normal or corrected-to-normal vision.

Design. Each participant was presented with 21 experimental trials together with 19 additional filler items. On the filler trials, one of the pictures was named in the spoken sentence. Thus, 82% of the 40 trials included a named picture; hence across trials participants may well have built up an expectation that one of the pictures would be named.

Materials were counterbalanced across the experimental trials for three groups of participants. Each participant received seven trials in the neutral, biasing, and shape competitor conditions. The same 19 fillers were used for all of the three groups. Trials were presented in the same random order to each participant.

Procedure. Each participant was tested individually and each was seated at a comfortable distance (with their eyes between 20 and 25 inches from the display) in front of the computer display. Throughout testing the participant wore an SMI EyeLink head-mounted eyetracker. Although viewing was binocular, the eyetracker sampled at 250 Hz from the right eye only. Participants were told that they should listen to the sentences carefully, that they could look at whatever they wanted to, but that they were not to take their eyes off the screen throughout the experiment. In other words, their only task was to listen to the spoken language while looking at the screen (cf. Altmann, 2004; Huettig & Altmann, 2005).

The onset of the presentation of the visual stimulus occurred 1 s before the onset of the spoken stimulus. The onset of the acoustic target word occurred on average 4 s after the onset of the spoken sentence, and thus the acoustic target word started to unfold on average 5 s after the onset of the visual stimulus. Between adjacent trials participants were shown a single dot located in the centre of the screen, which they were asked to fixate prior to a fixation cross appearing in this position (this procedure allowed the eyetracker to correct for drift). Participants would then press a response button for the next presentation. The termination of trials was preset and controlled by the experimental program, and thus participants could not terminate trials by themselves. The trial was automatically terminated after 9 s, which, typically, left 2 s after the end of the sentence. After every fourth trial, the eyetracker was recalibrated using a nine-point fixation stimulus. The EyeLink software automatically validates calibrations, and the experimenter, could, if required, repeat the calibration process if the validation was poor. Calibration took typically about 20 s. The entire experiment lasted approximately 20 min.

Data analyses. We examined certain time points as being of prime importance. The critical time points were: (a) at the onset of the critical word (henceforth the onset time point); (b) at its offset, (henceforth the offset time point); and (c) at 200 ms after its offset (henceforth the 200+ time point). Looks at the onset of the critical word are of interest in order to assess whether any biases in attention to any type of picture existed before information from the critical word became available. Looks at the offset time point reveal whether the unfolding of the critical word resulted in changes in overt attention. Given that it takes some time to program and initiate a saccadic eye movement (with estimates varying between 100 ms and 180 ms; see Altmann & Kamide, 2004, for review), fixation probabilities were also examined 200 ms after the offset of the acoustic target word (cf. Dahan & Tanenhaus, 2005). Eye movements initiated between 130 ms and 200 ms post-offset may reflect a "trigger" to move the eyes (and where to move them) that was received by the saccadic control mechanism post-offset, but it

is likely that the cognitive processes that caused that trigger took place at around word offset, if not just before. Despite this uncertainty, we can be sure that any divergence in the patterns emerging 200 ms post-offset are most likely due to later cognitive initiation than divergence emerging by the offset itself.

Our primary interest was whether more overt attention occurred to the critical pictures than to the unrelated distractors. To examine this, difference scores were calculated by subtracting the proportion of fixations to the distractor from the proportion of fixations to the target, and by subtracting the proportion of fixations to the distractors from the proportion of fixations to the competitor. Proportion of fixations to the distractors was averaged across the three distractor pictures. Difference scores reveal both the magnitude and direction of any tendency to favour one type of picture over another. Any positive difference reveals a bias of looks towards the critical picture, a negative difference reveals a bias to look towards the distractors, and difference scores close to zero reveal neither bias. We report below difference scores and their 95% confidence intervals (hence permitting statistical inference regarding the relationship between looks to the target/competitor object and looks to the distractors).

Results

At the *onset* time point there were no significant differences between looks to the target pictures and looks to the distractors in the neutral condition (mean difference score: 3.75; by participants, $p > .1$, upper 95% confidence interval (CI): -10.25, lower 95% CI: -2.75; by items, $p > .1$, upper 95% CI: 10.78, lower 95% CI: -3.73) and between looks to the shape competitors and looks to the distractors in the competitor condition (mean difference score: 1.38; by participants, $p > .1$, upper 95% CI: 6.79, lower 95% CI: -4.04; by items, $p > .1$, upper 95% CI: 9.29, lower 95% CI: -7.01). Therefore, there were no reliable biases in attention to any type of picture at the onset of the target word in the neutral and competitor conditions.[2] Table 1 reveals that there was a higher probability to fixate the target in the biasing condition at the onset of the critical word. This reliable bias (mean difference score: 20.92; by participants, $p < .001$, upper 95% CI: 28.92, lower 95% CI: 12.92; by items, $p < .001$, upper 95% CI: 28.26, lower 95% CI: 13.26) was

[2] Participants did not show significantly increased fixations to the visual competitor (e.g., the cable) at any point in time *before* hearing the target word. Note that Table 1 also shows the probability to fixate the types of pictures at the offset of the noun (“man” or “zookeeper”), the verb (“watched” or “worried”), and the adverb (“closely” or “greatly”) in the neutral or biasing phrases preceding the target word.

TABLE 1
Averaged probabilities of fixating a type of picture in Experiment 1

	Condition					
	Neutral		*Biasing*		*Competitor*	
Type of picture	*Target*	*Distractor*	*Target*	*Distractor*	*Shape competitor*	*Distractor*
p(fix) at prior noun ("man" or "zookeeper") offset	0.19	0.26	0.24	0.25	0.23	0.25
p(fix) at prior verb ("watched" or "worried") offset	0.21	0.26	0.34	0.21	0.26	0.24
p(fix) at prior adverb ("closely" or "greatly") offset	0.21	0.25	0.38***	0.20	0.25	0.25
p(fix) at target ("snake") onset	0.27	0.24	0.39***	0.19	0.25	0.24
p(fix) at target offset	0.50***	0.15	0.52***	0.14	0.33**	0.22
p(fix) at target offset + 200 ms	0.68***	0.09	0.62***	0.10	0.48***	0.17

*Difference score to distractors $p < .05$ for participants.
**Difference score to distractors $p < .01$ for participants.
***Difference score to distractors $p < .001$ for participants.

expected on the grounds that the prior biasing context ought to induce looks towards the relevant picture.

At the offset time point there were reliable biases in overt attention to the critical pictures in the neutral (mean difference score: 35.17; by participants, $p < .001$, upper 95% CI: 42.89, lower 95% CI: 27.44; by items, $p < .001$, upper 95% CI: 45.28, lower 95% CI: 24.91), biasing (mean difference score: 38.48; by participants, $p < .001$, upper 95% CI: 46.40, lower 95% CI: 30.52; by items, $p < .001$, upper 95% CI: 47.09, lower 95% CI: 29.86), and competitor conditions (mean difference score: 10.95; by participants, $p < .01$, upper 95% CI: 17.50, lower 95% CI: 4.71; by items, $p < .05$, upper 95% CI: 20.43, lower 95% CI: 1.48). Given the time it takes to initiate and program a saccadic eye movement means that the shifts towards the visual-shape competitors were initiated well before word offset. These effects are maintained (and indeed magnified) at the 200+ time point (see Table 1).

Finally, to examine in more detail the properties of this visual competitor effect, we correlated the eye movement data with the visual similarity ratings described earlier (i.e., the similarity of the depicted cable to the shape associated with the printed word "cable"). There was no statistically significant correlation between these ratings and the probability of fixating the visual competitor at either word offset or at word offset + 200 ms. However, for saccades launched towards the competitor during the acoustic lifetime of the target word (i.e., between word onset and word offset), the

subsequent fixation durations correlated significantly with the visual similarity ratings (Pearson correlation, two-tailed, $r = .55$, $p = .01$).

Supplementary graphical presentation. In addition to the statistical analyses, we plotted time-course graphs that illustrate the fixation probabilities to the various types of pictures over time. Note that the data are plotted in this way to aid visualization of participant performance—statistical analyses were carried out on the absolute proportion of trials on which the object of interest was fixated at a particular time point, *irrespective* of when that fixation was initiated. Figure 2 shows a time-course graph that illustrates the change in fixation probabilities at 20 ms intervals to the

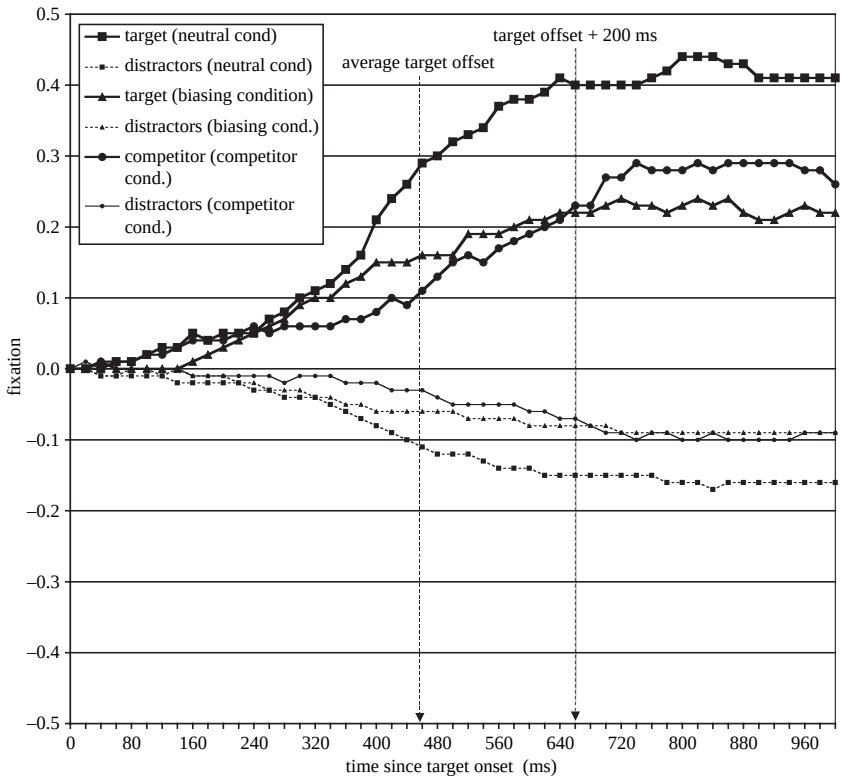

Figure 2. Proportion of trials with a fixation on the target in the neutral condition and the biasing condition, and on the shape competitor in the competitor condition (and averaged distractors of each condition). The curves are synchronized to the acoustic onset of the target word, and the x-axis shows time in milliseconds from this onset. The calculation excluded all movements prior to the acoustic onset, and thus negative values reflect moves away from objects that were already fixated at this onset; in effect, each data point reflects the proportion of trials with a fixation at that moment in time minus the proportion of trials with a fixation at the acoustic onset of the target word.

various types of pictures over the course of the average trial. In computing these values, each visual display was treated as being composed of four (virtual) quadrants, and eye position was categorized according to the currently fixated quadrant. *p*(targ) refers to the probability of fixating the target at a particular time point, *p*(comp) in this case refers to the probability of fixating the visual-shape competitor, and *p*(dist) refers to the averaged probability of fixating any distractor. The plots show the time course of fixations in the neutral, biasing, and competitor conditions, and individual curves relate to looks to the target in the neutral and in the biasing conditions, looks to the competitor in the competitor condition, and looks to the distractors in all of these conditions. Plots start from the acoustic onset of the critical word and cover the ensuing 1000 ms. This time region is of interest because it reflects the change in fixations that occurred both during and after the acoustic unfolding of the critical word.

In plotting the data, we follow Huettig and Altmann (2005) and plot only those data pertaining to fixations whose initial onsets occurred at or after the acoustic onset of the target word. Thus, the plots indicate *change* in fixation probability, with positive values indicating a net increase in the probability of fixation relative to the probability *at word onset*, and negative values indicating a net decrease in the probability of fixation relative to that probability. Consequently, the plots do not indicate any differences in the probabilities of fixation that might have existed at target word onset. These values are reported in Table 1.

The graphs in Figure 2 illustrate that participants shifted their overt attention to the target (e.g., snake) and the competitor (e.g., cable) during the critical time period very rapidly. The fixation probability curves shown suggest that *p*(comp), in the competitor condition, diverges from *p*(dist) very early. In other words, there was no observable delay in the time course of the shifts in overt attention to the shape competitor objects.

Discussion

To summarize the data: In the neutral condition no differences in overt attention to any particular type of picture were observed at the acoustic onset of the target word. At the later time points, however, attention was directed towards the corresponding target pictures. In contrast, in the biasing condition, the biasing sentential context gave rise to a substantial bias to look towards the target picture even before the target word had been presented. This bias was present in the data for all three time points. Finally, in the competitor condition, no differences in overt attention were found at the onset time point nor at earlier time points, but by the offset of the target word, there were reliable biases in overt attention towards the competitor

object (e.g., the cable), and the time spent fixating that object correlated significantly with the visual similarity ratings from the preliminary norming study. In other words, there was a robust shift in overt attention towards a picture of a shape competitor to the named target even though the competitor and target were conceptually unrelated. Importantly, however, there was no such shift towards the visual competitor before the target word acoustically unfolded.

The data from Experiment 1 reveal that when people listen to spoken language and are simultaneously presented with a number of visual objects, they direct spurious eye movements to objects that share only a visual-shape relationship with the concept activated by the spoken target word. This relationship is established very rapidly during spoken word processing—well before word offset—and appears to be one of the determinants of the time spent fixating the visually similar object. We return to the implication of these findings in the General Discussion.

The finding that sentential context biasing towards the target (snake) resulted in increased attention towards the target object (snake) even before the target word was heard is consistent with previous visual world findings that the mapping process between spoken language and visual objects occurs (at least) partially on the level of semantic/conceptual representations (cf. Cooper, 1974; Huettig & Altmann, 2005; see also the literature on anticipatory eye movements, e.g., Altmann & Kamide, 1999). Critically, there was no equivalent tendency to attend preferentially to the competitor object (the cable) in the biasing context. Only when the word "*snake*" was heard did attention towards this object increase. Consequently, whatever causes increased attention in the biasing condition towards the snake after hearing "*zookeeper*", but before hearing "*snake*", does *not* cause increased attention towards the cable in the competitor condition. The data thus rule out an explanation of this pretarget word bias in terms of "*zookeeper*" causing the activation of shape representations associated with snakes which are then matched against visual form information extracted directly from the image. Such an account would predict increased looks towards objects with similar visual forms, and no such increase was observed. We conclude that the bias to look towards the snake in the biasing condition is due to the existence of an episodic representation of the depicted snake that, being conceptually related to those conceptual representations associated with zookeepers, receives additional activation when "*zookeeper*" is encountered—thereby "attracting" attention back towards the snake.

These and prior data indicate that shifts in overt visual attention occur towards items related to words in the language when there is some featural match between the target specification accessed by the spoken word and the properties of the objects in the visual display. The shift, during the word "*snake*", towards the picture of the snake in the neutral and biasing

conditions of Experiment 1 reflects a full (or relatively full) featural match between the episodic representation of the snake and the target conceptual representation activated on hearing "*snake*". The shift towards the cable reflects a partial featural match between the episodic representation of the cable (including its visual form) and the target conceptual representation (including prototypical visual form) activated by "*snake*". We note that this partial featural match may, or may not, occur in the context of a phonological mismatch. If the episodic representation of the cable does not include its phonological form (i.e., the phonology associated with the word "cable"), then there is no phonological mismatch against "*snake*". Consequently, the data from Experiment 1 do not speak directly to how phonology might modulate the activation of the episodic representations that arise through inspection of the visual scene. There are, however, cases in which phonology might play a very crucial role in creating competitor objects (i.e., objects other than the intended target to which attention might be directed), and these cases are explored in Experiment 2. This second experiment further explores the conditions under which visual competitor effects arise, and asks, specifically, whether the contextual appropriateness of the target concept (cf. the snake concept in Experiment 1) modulates looks towards the visual competitor (cf. the cable). In effect, we ask how robust visual competitor effects are when there exists in the visual scene a contextually more appropriate object to which visual attention can be directed as the language unfolds.

EXPERIMENT 2

Experiment 2 was a variant of Experiment 1—in the neutral and biasing conditions, each display depicted an object related to the dominant meaning of the target word (a writing pen), an object related to the subordinate meaning of the target word (a cage-like enclosure), and two unrelated distractors. The biasing condition biased towards the pen-as-enclosure meaning. In the competitor conditions, the object related to the dominant meaning (the writing pen) was replaced by an object (a sewing needle) with similar visual shape as that associated with the dominant meaning. The context biased towards the pen-as-enclosure meaning.

Method

Materials. There were three experimental conditions: a *neutral condition*, a *biasing condition*, and a *competitor condition*. The design of the neutral and biasing conditions was as in Experiment 1. Figure 3a provides an example of the sort of display used in these conditions. Similar displays were used in the competitor condition (see Figure 3b), but the object corresponding to the dominant meaning (e.g., pen-writing implement) was replaced with the

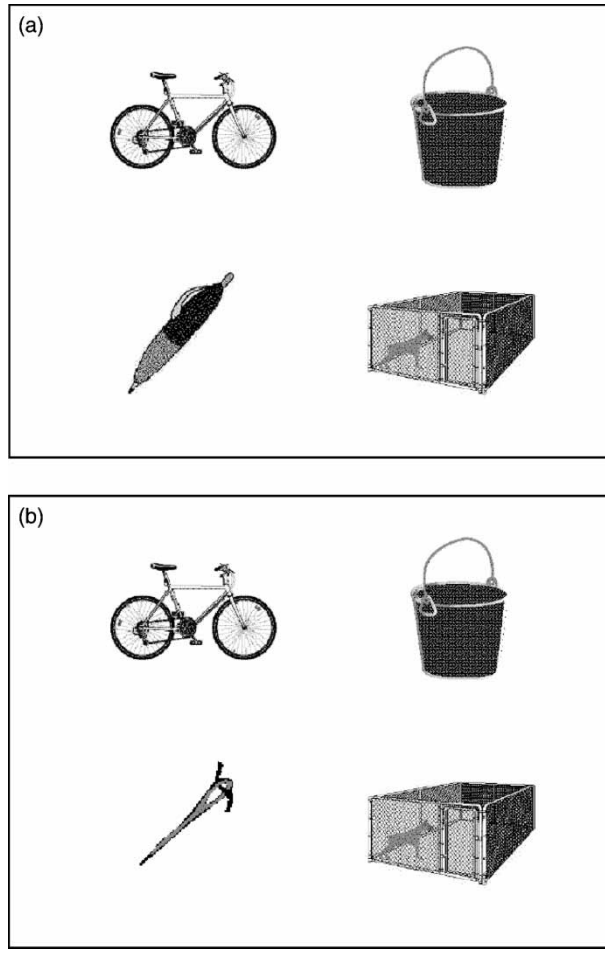

Figure 3. Examples of visual displays used in Experiment 2: (a) In the neutral and biasing conditions, and (b) in the competitor condition. Here the target homonym is "*pen*", and in both cases the pictures of the bicycle and bucket are distractors and the cage is the subordinate referent. In (a) the dominant referent is the pen-writing implement, and in (b) the needle is the competitor.

referent of a shape competitor (i.e., a needle). The visual stimuli were selected from commercially available ClipArt packages and presented in greyscale format. The selected shape competitors were jar (for battery-radio), robot (for boxer-fighter), pear (for bulb-light), teddy (for calf-cow), snowflake (for diamond-jewel), stage (for film-movie), planet (for heart-organ), torch (for horn-loud), picture (for letter-mail), needle (for pen-writing implement), hair (for plant-flower), ladle (for spade-shovel), bridge (for table-furniture), and handkerchief (for toast-bread). Naming agreement on all pictures was

collected from 69 participants. Responses were coded as intended, unintended, or "no response". Responses were coded as intended only when they exactly matched the intended name (e.g., "pen" was coded as valid but not "fountain pen"). "No response" was given in only 0.31% of trials. The intended response was given in 64% of trials. Unintended names were largely due to choosing a near-synonym ("mug" instead of "cup", or "Jupiter" instead of "planet"). Unintended names due to misidentification ("ball" instead of "planet") occurred on only 2.92% of trials.

The spoken sentence in the neutral condition did not bias either meaning of the homonym (up to the point when the homonym was heard): "*First, the man got ready quickly, but then he checked the pen and suspected that it was damaged.*" The sentences in the biasing and competitor conditions were identical, but were designed to bias interpretation towards the subordinate meaning: "*First, the welder locked up carefully, but then he checked the pen and suspected that it was damaged.*" These sentences were identical to those used in the neutral condition except for the biasing phrase ("*welder locked up carefully*"), which was replaced with a neutral phrase in the neutral condition ("*man got ready quickly*"). In the neutral and in the biasing condition, participants heard "*pen*" and saw a pen-writing implement, a pen-cage, and two unrelated distractors. In the competitor condition, participants heard "*pen*" and saw a needle (the shape competitor), a pen-cage, and two unrelated distractors. The pen-writing implement was not depicted in the competitor condition.

Preliminary norming studies

Norming study 1—rating word association. A word association task was carried out to establish the relative frequencies of the meanings of the homonyms. Participants were simply asked to write down the first word they thought of when reading each of the 15 homonyms. The relative meaning frequency was determined by the frequency of all responses across all participants (e.g., Nelson, McEvoy, Walling, & Wheeler, 1980; Twilley, Dixon, Taylor, & Clark, 1994). The assumption here is that participants produce associates to the lexical items (e.g., the response "money" after reading the word "bank") in proportion to the availability of the different meanings from the surface form of the word.

Twenty participants from the University of York student community took part. The selected homonyms and their relative meaning frequencies according to this norming study, the Twilley et al. (1994) norms, and the Nelson et al. (1980) norms are included in Table 2. Visual inspection of the data reveals that the homonyms were polarized because there was a large difference in the ratings for the dominant and subordinate meanings across all of the norms.

TABLE 2
The relative frequencies of the meanings of the homonyms used in Experiment 2 according to the present norming study, the Twilley et al. (1994) norms, and the Nelson et al. (1980) norms

	Current norms		*Twilley norms*		*Nelson norms*	
Homonym	*Dominant*	*Subordinate*	*Dominant*	*Subordinate*	*Dominant*	*Subordinate*
Battery	.90 (car)	.10 (hen)	.84	0	—	—
Boxer	.85 (fighter)	.10 (dog)	—	—	—	—
Bulb	.95 (light)	.05 (garden)	.86	.11	—	—
Calf	.85 (cow)	.15 (leg)	.81	.11	—	—
Diamond	1.0 (jewel)	0 (card)	.89	.01	.93	0
Film	.65 (movie)	.25 (photo)	.90	.02	—	—
Heart	.95 (lungs)	.05 (card)	—	—	—	—
Horn	.90 (car)	.05 (cow)	.77	.17	.91	.04
Letter	.95 (mail)	.05 (alphabet)	.68	.07	.91	.04
Pen	1.0 (pencil)	0 (pig)	.91	.04	.85	.09
Plant	.95 (green)	.05 (power)	.93	.02	—	—
Spade	.85 (garden)	.15 (card)	.66	.27	.48	.48
Table	1.0 (chair)	0 (figure)	—	—	—	—
Temple	.85 (god)	.15 (head)	.84	.02	.62	.31
Toast	.95 (jam)	.05 (wine)	.88	.09	.89	.09
Average	.91	.08	.83	.08	.80	.15

Norming study 2—rating strength of sentential context. A second norming study was also run to ensure that the linguistic contexts were appropriate. A different set of 24 participants from the University of York student community took part. Now participants were provided with a randomized list of the written experimental sentences up to the point where the critical word (the homonym) occurred in the neutral (e.g., "First, the man got ready quickly, but then he checked the pen ...") and the biasing/competitor (e.g., "First, the welder locked up carefully, but then he checked the pen ...") conditions.

They were asked to rate the particular strength and meaning bias of each sentence on a scale from -5 to $+5$, where -5 represented a strong bias towards the subordinate meaning (e.g., pen-cage), $+5$ a strong bias towards the dominant meaning (e.g., pen-writing implement), and zero represented neither bias. They were provided with an associate word of each meaning of the homonym to indicate which meaning was intended.

The average rating for the sentences in the biasing condition was -3.48 ($SD = 0.77$; upper bound of 95% confidence interval of mean: -3.04, lower bound: -3.90). Therefore the sentences in the biasing condition were biasing the subordinate meaning.

The average rating for the sentences in the neutral condition was 0.83 ($SD = 0.78$; upper bound of 95% confidence interval of mean: 1.27, lower

bound: 0.40). Therefore, there was a slight bias towards the dominant meaning for the sentences in the neutral condition. Indeed, it may be inherent in the strong frequency dominance effect that participants judge even ostensibly neutral sentences to be biasing slightly towards the dominant meaning.

Norming study 3—rating shape similarity. In order to determine the similarity of the shape of the dominant referent with the depicted objects, a final norming study was conducted. Ten participants were presented with the written (dominant meaning) of the homonym (e.g., pen-writing implement) and the pictures used in the corresponding displays. Participants were asked to judge how similar the typical physical shape of the dominant referent was to the physical shape of the depicted objects on a scale from 0 to 10 (0 representing: "absolutely no similarity in physical shape", 10 representing: "identical in physical shape").

The mean for the shape competitors was 4.63 ($SD = 2.41$), 0.71 ($SD = 0.55$) for the distractors, and 1.94 ($SD = 2.42$) for the subordinate referents. These differences in the shape similarity judgements were statistically significant, $F(2, 26) = 14.33$, $MSE = 3.92$, $p < .001$. Planned comparison revealed that there were no statistically significant differences between the distractors and the subordinate referents, $F(1, 13) = 3.59$, $MSE = 2.94$, $p > .05$. However, there were statistically significant differences between the shape competitors and the subordinate referents, $F(1, 13) = 9.26$, $MSE = 5.45$, $p < .01$, and the shape competitors and the distractors, $F(1, 13) = 32.0$, $MSE = 3.35$, $p < .01$. Therefore, the competitor pictures were judged to be significantly more similar in physical shape to the dominant meaning than to the other pictures.

Visual world study

Design. There were 15 experimental trials and 25 filler trials. A within-participants counterbalanced design was used across the three conditions. In counterbalancing Group A, each participant received five items in the neutral condition, five items in the biasing condition, and five items in the competitor condition. Assignment of these sets of five items were counterbalanced over two other groups (Groups B and C). The same 25 fillers were used for all of the three groups. The trials were presented in fixed random order as in the previous experiment.

Participants. Forty-eight participants from the University of York student community took part in this study. All were native speakers of British English and had either uncorrected vision or wore soft contact lenses or glasses.

Procedure. The procedure was the same as in Experiment 1.

Results

The results were analysed in the same way as in the previous experiment.

Neutral condition. Table 3 summarizes the fixation proportions of the current data. As summarized in Table 3, at the acoustic onset of the critical word, the probability to fixate the dominant referent, henceforth *p*(fix dom), the probability to fixate the subordinate referent, henceforth *p*(fix sub), and the probability to fixate the unrelated distractors, henceforth *p*(fix dist) were similar. There were no significant differences between looks to the dominant referents and looks to the distractors (mean difference score: 1.88; by participants, $p > .1$, upper 95% CI: 10.04, lower 95% CI: -6.29; by items, $p > .1$, upper 95% CI: 10.75, lower 95% CI: -6.62) and between looks to the subordinate referents and looks to the distractors (mean difference score: 5.63; by participants, $p > .1$, upper 95% CI: 14.03, lower 95% CI: -2.71; by items, $p > .1$, upper 95% CI: 14.34, lower 95% CI: -3.14). In other words, all types of pictures were treated as being equal in terms of the allocation of attention at this point.

At the acoustic offset of the target word, there were statistically significant biases in overt attention to the dominant referent (mean difference score: 18.13; by participants, $p < .001$, upper 95% CI: 26.90, lower 95% CI: 9.35; by items, $p < .05$, upper 95% CI: 32.38, lower 95% CI: 4.01), and to the subordinate referent (mean difference score: 9.79; by participants, $p < .01$, upper 95% CI: 18.23, lower 95% CI: 1.35; by items, $p < .05$, upper 95% CI: 18.66, lower 95% CI: 1.07). At the acoustic offset + 200 ms, statistically reliable biases in overt attention were observed to the dominant referent (mean difference score: 33.33; by participants, $p < .001$, upper 95% CI: 42.33, lower 95% CI: 24.34; by items, $p < .001$, upper 95% CI: 47.33, lower 95% CI: 19.20), and to the subordinate referent (mean difference score: 19.17; by participants, $p < .001$, upper 95% CI: 26.55, lower 95% CI: 11.79; by items, $p < .01$, upper 95% CI: 29.20, lower 95% CI: 6.70).

Biasing condition. Here the sentential context was designed to bias the subordinate referent of the homonym. Table 3 summarizes the fixation proportions. At the acoustic onset of the target word, there was no statistical difference in overt attention to the dominant referent relative to the distractors (mean difference score: 2.71; by participants, $p > .1$, upper 95% CI: 9.01, lower 95% CI: -3.60; by items, $p > .1$, upper 95% CI: 13.78, lower 95% CI: -8.32). However, there was a reliable bias in overt attention towards the subordinate referent relative to the distractors (mean difference score: 28.52; by participants, $p < .001$, upper 95% CI: 38.58, lower 95% CI: 18.50; by items, $p < .001$, upper 95% CI: 36.79, lower 95% CI: 20.41). The difference in overt attention towards the subordinate referent relative to that towards the

TABLE 3
The probability of fixating each type of picture at the acoustic onset, offset, and 200 ms after the offset of the target words (averaged for participants and items) in the neutral, biasing, and competitor conditions in Experiment 2

	Type of picture		
Time point	*Dominant referent*	*Subordinate referent*	*Distractor*
Neutral condition			
p(fix) at onset	0.24	0.28	0.22
p(fix) at offset	0.34***	0.26**	0.16
p(fix) at offset + 200 ms[a]	0.43***	0.29***	0.10
Biasing condition			
p(fix) at onset	0.19	0.45***	0.16
p(fix) at offset	0.23**	0.45***	0.13
p(fix) at offset + 200 ms[a]	0.27***	0.46***	0.09
	Shape competitor	*Subordinate referent*	*Distractor*
Competitor condition			
p(fix) at onset	0.20	0.49***	0.15
p(fix) at offset	0.22**	0.50***	0.13
p(fix) at offset + 200 ms	0.26***	0.50***	0.10

The biasing context refers to sentences that biased the subordinate meaning of the homonym.
*Difference score to distractors $p < .05$ for participants.
**Difference score to distractors $p < .01$ for participants.
***Difference score to distractors $p < .001$ for participants.

distractors is maintained at the later time points. At the offset of the target word there was more overt attention to the (contextually inappropriate) dominant referent relative to the unrelated distractors (mean difference score: 9.79; by participants, $p < .01$, upper 95% CI: 16.09, lower 95% CI: 3.50; by items, $p = .07$, upper 95% CI: 20.55, lower 95% CI: -1.08). Statistically reliable biases were found at the later offset + 200 ms time point (mean difference score: 17.91; by participants, $p < .001$, upper 95% CI: 24.85, lower 95% CI: 10.98; by items, $p < .01$, upper 95% CI: 29.16, lower 95% CI: 6.70). Thus, the sentential context biasing the subordinate meaning did not prevent eventual increased overt attention to the dominant referent.

Competitor condition. In this condition, the sentential context was designed to bias the subordinate meaning of the homonym. Importantly, however, the visual referent of the dominant meaning (e.g., pen-writing implement) was replaced with a shape competitor (e.g., a needle). Table 3 summarizes the fixation proportions. There was a strong bias in attention towards the subordinate referent (mean difference score: 34.17; by

participants, $p < .001$, upper 95% CI: 42.95, lower 95% CI: 25.39; by items, $p < .001$, upper 95% CI: 45.45, lower 95% CI: 22.82), but no significant bias to fixate the competitor at the onset time point (mean difference score: 5.83; by participants, $p > .05$, upper 95% CI: 12.25, lower 95% CI: -0.59; by items, $p > .1$, upper 95% CI: 13.83, lower 95% CI: -2.09). The difference in overt attention towards the subordinate referent relative to that towards the distractors is maintained at the later time points. There was a significant bias in attention towards the shape competitor, relative to the distractors, at word offset (mean difference score: 8.96; by participants, $p < .05$, upper 95% CI: 16.77, lower 95% CI: 1.14; by items, $p < .05$, upper 95% CI: 16.89, lower 95% CI: 1.24). More overt attention was also directed towards the shape competitor relative to the unrelated distractors at the offset + 200 ms (mean difference score: 15.63; by participants, $p < .001$, upper 95% CI: 22.45, lower 95% CI: 8.80; by items, $p < .05$, upper 95% CI: 27.62, lower 95% CI: 3.84) time points.

Supplementary graphical presentation. Figure 4a shows the time-course graph in the neutral condition. As before, the time-course graph shows the change in fixation probabilities relative to the probability of fixation at target word onset—that is, only fixations initiated from that point onwards are included in the graph. *p*(fix sub) stayed at a similar level throughout the total time window. However, performance associated with the subordinate referent is unlike that associated with an unrelated distractor. It can be seen that there was a gradual decrease in *p*(fix dist) during the acoustic lifetime of the critical word. If the subordinate referent had been treated like an unrelated distractor, then a similar effect would have been expected. In other words, although the fixation probabilities to the subordinate referent did not rise in this condition, it was nevertheless *privileged* in terms of allocation of attention compared to the unrelated distractors. The fact that the *p*(fix sub) did not show an increase of greater magnitude is most likely due to the fact that the dominant referent attracted the most attention. In other words, the subordinate referent was competing for attention with the dominant referent. This shows very clearly how the probability to fixate a particular visual referent is necessarily determined (in part) by what other referents compete for attention in the same display.

Figure 4b shows the change in fixation probabilities relative to the probability of fixation at target word onset in the biasing condition. Figure 4c shows this time course in the competitor condition. From Figure 4c one can see that looks towards the visual form competitor (e.g., the needle) do rise relative to looks towards the distractors. Recall that the graph takes into account (and eliminates) the bias to fixate the cage at the onset of the target word. Thus, in absolute terms, there were more fixations on the cage than on the needle throughout (see Table 3c). However, in terms of the *change* in fixation

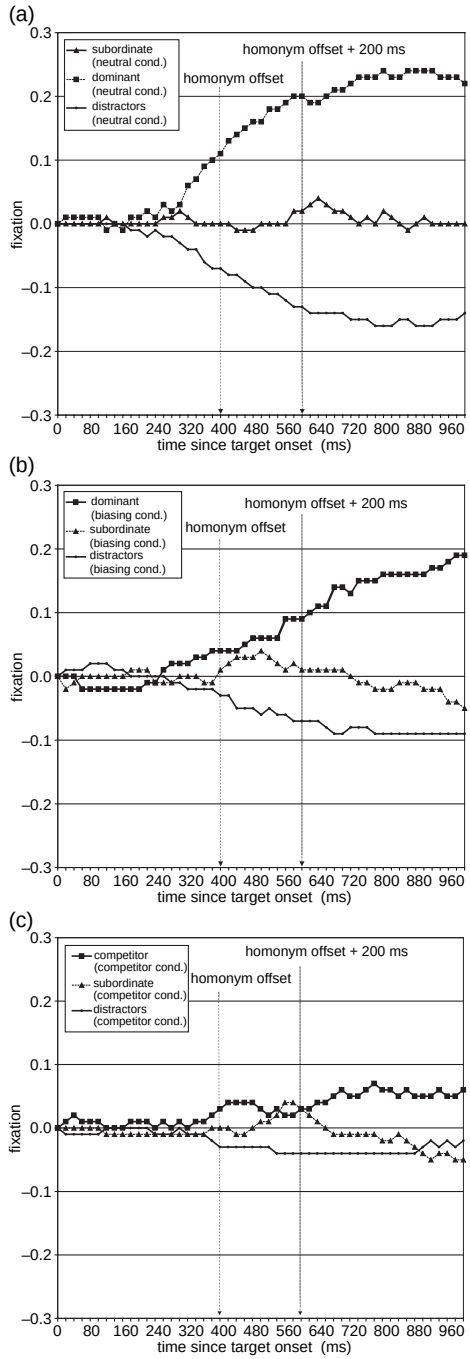

Figure 4 (*Continued*)

probability, Figure 4c shows how fixation probabilities on the cage peak at around the offset + 200ms and then drop down somewhat, whilst fixation probabilities on the needle steadily increase from around word offset onwards.

Discussion

The key finding in the competitor condition of Experiment 2 was that attention was directed towards a visual referent that was similar in shape to the dominant referent of a heard homonym even though (a) there was a picture of the subordinate referent present, and (b) the linguistic context biased the subordinate meaning of the homonym. Looks to the shape competitor were more likely than were looks to an unrelated distractor. The data thus provide evidence for the activation of the inappropriate dominant meaning of the word "*pen*" even though no writing implement was present in the display. This suggests that the perceptual (visual-shape) representations of the contextually inappropriate dominant referent were accessed even though the contextually more appropriate subordinate referent was depicted in the scene. Had there been no more looks towards the sewing needle than towards the distractors, we would have had to conclude that visual competitor effects are modulated by the contextual appropriateness of the target concept. In the event, we cannot completely rule out such modulation, and indeed, compared to the visual competitor effect found in Experiment 1, the effect found here appears relatively attenuated. However, this may in part be due to the copresence of a visual referent that more completely matches the featural specification of the target word "*pen*"—in this respect, it is not particularly surprising that there was an attenuated competitor effect relative to Experiment 1. Crucially, even if there is some modulation by the context, it was not enough, in Experiment 2, to eliminate competitor effects entirely. Visual competitor effects are thus robust in the face of contextually more appropriate objects to which attention is directed by the unfolding language (see the General Discussion).

In respect of the neutral and biasing conditions, the data are entirely compatible with current accounts of lexical access—in a neutral context, there were more looks after "*pen*" towards the dominant referent (the pen-writing implement) than towards the subordinate referent, but this effect was mediated, in the biasing condition, by the context which favoured looks towards the subordinate referent (the pen-enclosure). And in this latter case, although there was a consistently greater probability of fixation on the

Figure 4. Proportion of trials with a fixation on: (a) the dominant referent, the subordinate referent, and the distractors in the neutral conditions; (b) the dominant referent, the subordinate referent, and the distractors in the biasing conditions; (c) the shape competitor, the subordinate referent, and the distractors in the competitor condition, in Experiment 2. The cues are synchronized to the acoustic onset of the target word, and the x-axis shows time in milliseconds from this onset.

subordinate referent, only fixations on the dominant referent *increased* during and beyond the target word "*pen*"—in effect the phonological information, and the activation of the associated conceptual representation, modulated looks towards the dominant meaning only (although all this means is that looks to the subordinate meaning were at ceiling because of the prior sentential bias).

GENERAL DISCUSSION

By systematically controlling the relationship between the objects in the visual displays and the unfolding speech signal, it has been possible to control the extent to which semantic/visual features associated with the visual objects match those associated with the concepts activated as the speech unfolds. Experiment 1 was concerned with the manner in which information about the shape of objects is activated from the spoken language. It was found that on hearing "*snake*" participants shifted overt attention immediately towards a picture of a conceptually unrelated object that has a similar global shape: a cable. The finding that more looks were directed towards the cable, upon hearing "*snake*", than towards any of the distractors suggests that hearing "*snake*" activated visual-shape information that overlapped with the visual-shape of the visually concurrent cable (cf. Dahan & Tanenhaus, 2005).

These data are consistent with the notion that language-mediated eye movements can be directed to objects that share some characteristics, but not all, with the target specification determined by the unfolding word (Huettig & Altmann, 2004). Huettig and Altmann (2005) found that language-mediated eye movements are a sensitive index of the overlap between the conceptual information conveyed by spoken words and the conceptual representations of the concurrent visual objects. Hearing "*piano*" causes us to attend to a trumpet because of the conceptual overlap between pianos and trumpets; activation of these common conceptual features by the word "*piano*" attracts attention to whatever in the visual field shares these conceptual features (see Huettig & Altmann, 2004). These conceptual competitor effects are predicated on the fact that participants viewed the objects in the visual display for approximately 5 s before hearing the target word—sufficient time in which to recognize the objects for what they were. This raises the obvious question: Why, when we hear "*snake*", should we move our eyes to the cable when, evidently, we know that what we want is an animal, and that what we will get is a conductor for transmitting electrical power? At least in the "*piano*"-trumpet case the two were conceptually related; the same cannot be said for cables and snakes.

The present data suggest that the answer to this is that participants shift their attention towards the visual object in the display that best matches the conceptual *and* perceptual specification of the concept activated by the spoken word. In other words, the data suggest that the best matching object

in the visual field is fixated even if this object has little in common with the target concept activated by the spoken target word. An overlap in characteristics such as an object's shape will result in shifts of overt attention to the visual object in the array that best matches the specification of the concepts accessed by the spoken words (cf. Huettig & Altmann, 2004, 2005). Indeed, on the assumption that experience of the events in which objects can participate (with us or with each other) guides concept formation (cf. McRae, Ferretti, & Amyote, 1997; Nelson, 1996), what we have hitherto termed "conceptual" and "perceptual" may not be easily separated in cases where perceptual form is inextricably bound to those experiences that have a perceptual (i.e., some sensorimotoric) basis.

In addition, note that the shape competitor effects are unlikely to be limited to the task employed here. Dahan and Tanenhaus (2005) demonstrated similar visual form competitor effects when participants were required to engage in an explicit physical task (moving the objects mentioned in spoken sentences using a computer mouse). Similarly, Yee and Sedivy (2006), using a task in which participants had to touch one of the displayed objects on a computer screen, observed similar semantic effects to those obtained by Huettig and Altmann (2005). In Experiment 1, we used a variant of the paradigm in which, on experimental trials in the competitor condition, the entity mentioned in the spoken sentence was not present in the visual display. But the two types of competitor effects found here (visual-shape and semantic) have also been observed when targets have been present (visual-shape competitor effects, Dahan & Tanenhaus, 2005; semantic competitor effects, Yee & Sedivy, 2006). Moreover, Huettig and Altmann (2005) compared target-present and target-absent conditions and have found, other than the tendency for fixations to targets to dominate when targets are present, similar results across these conditions. Thus, these competitor effects are not limited to certain specific goal-directed task demands.[3]

However, our data do more than attest to the generalizability of the Dahan and Tanenhaus (2005) demonstration of visual competitor effects; the fact that a biasing context can cause increased looks towards an object in a concurrent scene, but *not* towards a visual competitor of such an object,

[3] The visual form and semantic competitor effects also rule out certain task-specific strategies. It rules out the possibility that the mapping process is based merely on phonological information associated with the visual objects. In other words, it could not be the case that participants are covertly naming objects in the visual field, and then using the match between these names and the unfolding speech stream to guide eye movements towards whichever phonological matches are so obtained. Moreover, studies such as Salverda, Dahan, and McQueen (2003) have shown how manipulations of fine-grained speech detail, while keeping the display constant, modulate overt attention. Such effects can therefore not be attributed to a prenaming strategy.

constrains the range of explanations for how the eyes are directed, by the unfolding speech, towards objects in the visual field. We suggest that it is not the case that a spoken word activates a target concept whose visual features initiate some form of visual search for corresponding features in the concurrent scene (cf. Dahan & Tanenhaus, 2005). Rather, we suggest that our data are more compatible with an account in which a spoken word *re*activates (or increases activation of) a target concept previously activated by the visual experience of the corresponding object in the scene. It is not the case that a visual referent is *found* in response to a "top-down" target representation (as may be the case in standard visual search studies) activated from the spoken input. It may be worth spelling out a step-by-step description of task performance in our experiments. On display onset participants start to view the four objects. This causes picture-derived activations of representations (including visual form representations) that are tied to their spatial location in the display. Note in this regard, that participants had viewed the visual display with four objects for approximately 5 s before hearing the target word, and thus had sufficient time to recognize the objects for what they were. As participants hear the spoken language input language-derived representations (including visual form representations) are activated. Overlap between the picture-derived representations and the language-derived representations in turn causes eye movements to the competitor objects in the display. Thus, the spoken input *selects* from amongst the target representations activated from the *prior visual input*.[4]

More importantly, we found that a context that biases looks towards an object does not bias looks towards a visual competitor for that object—only when a word is encountered whose associated conceptual representations shares semantic/visual features with the target representations activated from the scene do the eyes move towards whatever in the scene activated those representations. Our data from Experiment 2 suggest, moreover, that conceptual matches of this kind are limited only to the extent that the conceptual representations activated by the spoken word might themselves be limited—thus, even though the context may favour one meaning of a homonym over another, so long as the other is activated, and so long as there are objects in the visual scene whose conceptual representations share features (even partially) with those activated representations, the eyes will be directed towards those objects (and even when some of those objects are better "fits" than others).

[4] 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.

Our results thus suggest that visual form competition during language-mediated attention is primarily lexically driven and not modulated by contextual appropriateness. Experiment 1 showed that when hearing about zookeepers, participants do not look at objects (a cable) that are visually similar to a snake (which is semantically related to zookeepers) even though they look at the snake itself when it is present in the display. However, as the word "*snake*" acoustically unfolds, participants shift overt attention to the cable. Thus, in Experiment 1, only hearing "*snake*" but not a contextual bias towards snakes compelled the visual form effect. It could be argued that the contextual bias in Experiment 1 was not sufficiently strong to induce such an effect. Experiment 2, however, presents even greater evidence for the primacy of lexical input. On hearing about welders, and with a pen-enclosure in the display, participants do not look at a picture depicting the alternative meaning of pen (nor the visual competitor of the alternative meaning). Only on hearing "*pen*" did participants look at the contextually inappropriate meaning of pen (or an object that shares some visual similarity with the contextual inappropriate meaning of pen).

Though not ruling out modulation by linguistic context, the findings presented here point towards the importance of lexical input for visual form competition during language-mediated attention.[5] It is interesting to note the points of contact between the present findings and those more commonly discussed in the context of models of lexical ambiguity resolution. In this regard, the close similarities between the present eye-movement effects and other effects found in the psycholinguistic literature are quite striking. The same key factors that have been found to influence lexical ambiguity resolution during reading appear to have been responsible for overt shifts in attention in Experiment 2. For instance, in the neutral context, the dominant referent received more overt attention than did the subordinate referent when the homonym was encountered. This finding accords well with effects concerning the relative frequencies of the alternative meanings of ambiguous words found in eyetracking experiments concerning reading (cf. Duffy et al., 1988). Similarly, the data support the notion that a sentential context biasing the subordinate meaning of the homonym cannot prevent the activation of the unintended and inappropriate dominant meaning (cf. Onifer & Swinney, 1981). The present visual world study thus also provides converging evidence for conclusions derived from other psycholinguistic methods.

[5] Some of the data reported here (e.g., the absence of modulation of the visual form effect by contextual appropriateness) may appear inconsistent with some embodied theories of cognition. In this regard, we would like to emphasize that the present research was not conducted to evaluate theories of conceptual representation. Future work, however, could usefully be directed at this issue.

Finally, the present data clearly show that visual-shape information can be activated very rapidly and well before word offset. Why otherwise would our participants have shifted overt attention towards visually similar objects? This raises the question why, using the lexical decision task, it has been so difficult to find clear evidence for perceptual priming (cf. Pecher et al., 1998) or why it was found to be delayed (cf. Moss et al., 1997). We can only speculate here. However, there is an important methodological distinction between the effects we, and others, have found within the visual world paradigm, and the cross-modal priming effects reported by others. Critically, in cross-modal priming, the auditory prime *precedes* the visual (orthographic) target. Thus, any visual properties associated with the conceptual representation activated by the spoken word could in principle become activated before the onset of the visual target. But looks to a visually or semantically related object in the visual scene are not the equivalent of facilitatory priming of the visual target in cross-modal priming. This is because in the visual world paradigm, the visual target *precedes* the auditory signal—visual properties associated with the conceptual representations ordinarily activated by the spoken word are *not* activated before the onset of the visual objects; those visual properties have already been activated, precisely because those visual objects precede the spoken word. The implication of this for how to interpret time-course information depicted in, for example, Figures 2 and 4 is that these graphs do not necessarily indicate the time course of activation of visual feature information associated with the unfolding word. Rather, they indicate the time course of the process by which this information modulates the activation of the episodic representations themselves activated, earlier, by the visual scene.

In sum, the present data indicate that when people listen to spoken language and are simultaneously presented with a number of visual objects, perceptual synergies between the concepts activated by spoken words and visual objects can mediate visual attention. Here we have shown that the mapping between language and visual input can be mediated by shape relations. We conclude that this visual form effect is primarily lexically driven, and that the fit between language-derived representations and the picture-derived representations can drive the eyes towards appropriate, or inappropriate, objects in the visual environment. Thus, such activation can under certain circumstances (e.g., during the processing of dominant meanings of homonyms) constrain the direction of visual attention even when that direction is clearly contextually inappropriate.

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*, 419–439.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm". *Cognition*, *93*, B79–B87.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision and action* (pp. 347–386). Hove, UK: Psychology Press.

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*(1), 84–107.

Cree, G. S., & McRae, K. (2003). 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*, *132*, 163–201.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*, 317–367.

Dahan, D., & Tanenhaus, M. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin and Review*, *12*(3), 453–459.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*, 429–446.

Flores d'Arcais, G. B., Schreuder, R., & Glazenborg, G. (1985). Semantic activation during recognition of referential words. *Psychological Research*, *47*, 39–49.

Henderson, J. M., & Ferreira, F. (2004). *The interface of language, vision and action*. Hove, UK: Psychology Press.

Huettig, 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 on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 187–207). New York: Psychology Press.

Huettig, 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.

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. M. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica*, *121*, 65–80.

Kellenbach, M. L., Wijers, A. A., & Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: Event-related potential evidence for perceptual semantic priming. *Cognitive Brain Research*, *10*, 67–75.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211–240.

McDonald, S. A. (2000). *Environmental determinants of lexical processing effort*. Unpublished doctoral dissertation, University of Edinburgh, Scotland. Retrieved December 10, 2004, from http://www.inf.ed.ac.uk/publications/thesis/online/IP000007.pdf

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, *12*, 137–176.

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.

Nelson, D. L., McEvoy, C. L., Walling, J. R., & Wheeler, J. W., Jr. (1980). The University of South Florida homograph norms. *Behaviour Research Methods and Instrumentation*, *12*, 16–37.

Nelson, K. (1996). *Language in cognitive development: The emergence of the mediated mind*. Cambridge, UK: Cambridge University Press.

Onifer, W., & Swinney, D. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency-of-meaning and contextual bias. *Memory and Cognition*, *9*, 225–236.

Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. W. (1998). Does pizza prime coin? Perceptual priming in lexical decision and pronunciation. *Journal of Memory and Language*, *38*, 401–418.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.

Schreuder, R., Flores D'Arcais, G. B., & Glazenborg, G. (1984). Effects of perceptual and conceptual similarity in semantic priming. *Psychological Research*, *45*, 339–354.

Simpson, G. B. (1994). Context and the processing of ambiguous words. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 359–374). New York: Academic Press.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Behaviour*, *18*, 645–659.

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

Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of meaning frequency for 566 homographs. *Memory and Cognition*, *22*, 111–126.

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), 81–92.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1–14.

Manuscript received February 2006
Manuscript accepted November 2006
First published online March 2007