Journal of Experimental Psychology: Human Perception and Performance

Attentional Capture of Objects Referred to by Spoken Language

Anne Pier Salverda, and Gerry T. M. Altmann

Online First Publication, April 25, 2011. doi: 10.1037/a0023101

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/11/$12.00 DOI: 10.1037/a0023101

Attentional Capture of Objects Referred to by Spoken Language

Anne Pier Salverda
University of Rochester

Gerry T. M. Altmann
University of York, UK

Participants saw a small number of objects in a visual display and performed a visual detection or visual-discrimination task in the context of task-irrelevant spoken distractors. In each experiment, a visual cue was presented 400 ms after the onset of a spoken word. In experiments 1 and 2, the cue was an isoluminant color change and participants generated an eye movement to the target object. In experiment 1, responses were slower when the spoken word referred to the distractor object than when it referred to the target object. In experiment 2, responses were slower when the spoken word referred to a distractor object than when it referred to an object not in the display. In experiment 3, the cue was a small shift in location of the target object and participants indicated the direction of the shift. Responses were slowest when the word referred to the distractor object, faster when the word did not have a referent, and fastest when the word referred to the target object. Taken together, the results demonstrate that referents of spoken words capture attention.

Keywords: visual attention, attentional capture, eye movements, lexical processing, visual-world paradigm

To facilitate the processing of visual information, attention is usually directed at a restricted part of the visual field. By focusing attention on a particular region of the visual field, processing of information at and around that location is enhanced. Theories of visual attention (see Egeth & Yantis, 1997, for a review) are concerned with how visual attention is deployed within the visual system, and traditionally make a distinction between two types of attentional control (also see Yantis, 2000). Attention is goal-driven (*endogenous*) when it is intentional and a result of deliberate strategies on the part of the observer to achieve a particular goal. For instance, when searching for a green apple, eye movements may be directed to green objects. Attention is stimulus-driven (*exogenous*) when it is controlled by a salient stimulus property, independent of the observer's goals and intentions. For instance, attention can be captured by the sudden appearance of a new object (Yantis & Jonides, 1984; Jonides & Yantis, 1988).

Theories of visual attention typically focus on the role of visual information in attentional control. However, in our interactions with the world in everyday life, visual processing is frequently informed by information derived from processing in other cognitive domains. For instance, when a speaker refers to an object in the immediate visual environment, a listener will usually attend to that object. If the visual system gives priority to visual information to determine the allocation of visual attention, one would not expect linguistic information (e.g., spoken words) to interfere with visual attention, in particular when carrying out a visual task where the linguistic information is task-irrelevant. Here, we question this idea, and present data to suggest that task-irrelevant linguistic information can influence visual attention. We thus follow James (1890) in considering the idea that words capture attention.

A growing body of experimental work is concerned with the interaction of visual and linguistic information. Studies using the "visual-world" paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; see Tanenhaus, 2007, for a review) have demonstrated that people readily and rapidly integrate visual and linguistic information when processing spoken instructions in the context of a task-relevant visual world. In these experiments, eye movements to referents of linguistic expressions are initiated with a surprisingly short delay relative to the processing of relevant information in the speech signal. For instance, Allopenna, Magnuson, & Tanenhaus (1998) presented participants with visual displays including four objects and spoken instructions to manipulate one of the objects, such as "Click on the beaker." They found that the odds of fixating the target object over the distractor objects (whose names were phonologically dissimilar from the name of the target object) increased as soon as 200 ms after the onset of the spoken word "beaker." This is well before the offset of the spoken word, which was approximately 400 ms in duration. Given an estimate of 150 to 200 ms for programming and executing an eye movement (Hallett, 1986; and see Altmann & Kamide, 2004, but cf. Altmann, in press, who describes circumstances in which language-mediation of oculomotor control may occur in as little as 100 ms following the onset of a word), it appears that participants very rapidly focused their attention on the target object, upon hearing the initial sounds of the target word.

Eye movements to visual referents in visual-world studies may reflect the use of task-relevant strategies by the observer. That is, a participant may integrate visual and linguistic information in the

Anne Pier Salverda, University of Rochester; and Gerry T. M. Altmann, University of York, UK.

This research was supported by awards to G.T.M.A. from the Medical Research Council (G0000224) and the Wellcome Trust (076702/Z/05/Z) and by NIH grant DC005071 to Michael K. Tanenhaus. We thank Tom Covey and Dana Subik for their assistance in stimulus preparation and data collection, Alice Cruickshank and Mike Tanenhaus for helpful comments, and Jan Theeuwes for helpful discussions.

Correspondence concerning this article should be addressed to Anne Pier Salverda, Department of Brain and Cognitive Sciences, University of Rochester, Meliora Hall, Box 270268, Rochester, NY 14627-0268. E-mail: asalverda@bcs.rochester.edu

visual-world paradigm specifically to be able to carry out the experimental task. Alternatively, the very speed with which participants integrate visual and linguistic information (as evidenced by rapid language-mediated shifts in visual attention observed in visual-world studies) may indicate that this process proceeds automatically, that is, independently of the task and hence without strategic control.[1] This would of course not imply that strategic, task-based, and/or goal-based processes do not affect eye movements in visual-world studies (see also Salverda, Brown, & Tanenhaus, in press, for an account of eye movements in visual-world studies that focuses on goal-based effects). Rather, the idea is that visual-linguistic integration proceeds automatically, and that this process influences eye movements in visual-world studies.

The notion that language-mediated eye movements may be automatic receives further support from studies showing that listeners fixate objects which are only related to the objects that the unfolding language refers to (e.g., saccades are launched to a piano on hearing "trumpet"; Huettig & Altmann, 2005; Yee & Sedivy, 2006; cf. Moores, Laiti, & Chelazzi, 2003; Meyer, Belke, Telling, and Humphreys, 2007, and Belke, Humphreys, Watson, Meyer, & Telling, 2008, for related findings in visual search). Moreover, recent studies show that language-mediated eye movements occur even when the scene to which the language refers has been *removed* prior to the unfolding of the language (Altmann, 2004; Altmann, in press; Altmann & Kamide, 2009). Accounts of the language-mediated control of eye movements, when the scene is concurrent or absent, or when the language refers only to related objects, propose that such eye movements come about through spreading activation from the language-mediated activation of conceptual structures to the overlapping conceptual structures previously activated by the contents of the visual scene—this reactivation of the "episodic traces" associated with objects in the concurrent or prior scene causes eye movements back toward the locations associated with those traces (Altmann & Kamide, 2007; cf. Richardson & Spivey, 2000).

The idea that visual attention can be influenced by concurrent cognitive representations, independently of task, has been explored recently in work showing that visual attention can be guided by task-irrelevant contents of working memory (see Soto, Hodsoll, Rotshtein, & Humphreys, 2008, for a review). Typically, these studies use a dual-task procedure including a memory task. For instance, in experiment 4 of a study by Soto, Heinke, Humphreys, & Blanco (2005), a trial started with the presentation of a colored shape which the participant was required to memorize (each trial concluded with a memory test for this prime stimulus). Subsequently, a new visual display appeared consisting of four colored shapes, each containing a vertical line. One of the lines was slightly tilted to the left or right. The participant's task was to report the orientation of the tilted line. In some trials, one of the four colored shapes matched the prime. In these trials, participants were slower to detect the orientation of the tilted line, compared to trials on which none of the colored shapes matched the prime. This finding suggests that attention was oriented automatically toward the shape that matched the shape held in working memory, and that this affected performance on the discrimination task. Crucially, a colored shape that matched the prime never contained the tilted line. The information held in working memory thus disrupted performance in the discrimination task, even though it was not relevant for that task. This suggests that guidance of visual attention by the contents of working memory is an automatic process, that is, a process that is not the result of a deliberate strategy.

In a closely related study that is particularly relevant for the current experiments, Soto and Humphreys (2007) compared the influence of visual and verbal primes on performance in a visual discrimination task much like the one used in experiment 4 of Soto et al. (2005). Here, at the beginning of a trial, participants saw either a visual prime (e.g., a picture of a red square) or a verbal prime (e.g., the words "red square") and were required to verbalize the prime. Interference effects were obtained for visual as well as linguistic primes. Interestingly, these effects appeared to be of the same magnitude. This suggests that memorized linguistic primes are equally effective as memorized visual primes in guiding visual attention to visual stimuli that match the prime. Interestingly, interference effects for both types of primes were also found when the discrimination task was not followed by a memory task, that is, when participants were required to verbalize but not memorize the prime—although under those conditions, performance appeared to be less disrupted by the prime. Taken together, the results of Soto and Humphreys' study suggest that visual attention is guided by information in working memory, and that, where available, verbally produced information is automatically encoded into working memory (cf. Meyer et al., 2007, who showed that a visual prime held in working memory influenced the allocation of attention toward semantically and phonologically related visual objects during a subsequent visual search task).

Soto and Humphreys' (2007) study bears a close relationship to the current study, which examines the influence of spoken words on the allocation of visual attention to referents of those spoken words. However, there are several important differences between our study and that of Soto and Humphreys (and other, related studies). Although effects in Soto and Humphreys' study were mediated by participants engaging in speech production (i.e., verbalization of a prime shape), effects observed in the studies we describe below are mediated by speech comprehension. Moreover, although Soto and Humphreys found influences of verbalization on *subsequent* visual search, in the current study, we explore more immediate influences of speech comprehension on visual attention. In each of three experiments, participants had to respond to a visual event that generally occurred during the acoustic lifetime of an unfolding spoken word. Most crucially, we examined the influence of referential spoken language on processing of these visual events under conditions where the linguistic information was purely incidental to the participant's task. In contrast, in Soto and Humphreys' and related studies (e.g., Downing, 2000; Moores & Maxwell, 2008; Olivers, Meijer, & Theeuwes, 2006; Soto et al., 2005), an influence of information in working memory on visual attention was obtained only under conditions where the experimental task required processing of the prime stimulus (e.g., producing a short phrase describing the prime, holding the prime in memory, or otherwise attending to the prime) immediately before performing the subsequent visual search or discrimination task. Thus, even though the prime stimulus was not relevant for the visual task, it *was* relevant, as far as the participants were concerned, for the experimental task they were required to perform.

[1] For ease of exposition, we use the term "automatic" for task-independent attentional control.

Therefore, the automatic guidance of visual attention from working memory arguably took place only as a result of the mandatory processing of the prior information. It is thus unclear whether a linguistic prime would affect performance in a subsequent discrimination task if the prime were truly task-irrelevant, that is, when processing the linguistic prime is not required for carrying out any aspect of the experimental task. In our experiments, participants were not required to process the linguistic information—in fact they are encouraged to ignore the spoken words.

To summarize, although there is agreement in the literature that the allocation of visual attention can be influenced by language-related information, the evidence to-date is less clear about whether such influences obtain when the processing of the linguistic information is not mandatory, is strategically counterproductive, and when the linguistic information unfolds across the same moments in time as does the visual information to which attention must be directed. The primary goal of the studies described below is to establish whether the integration of visual and linguistic information proceeds automatically. If it does, we would predict that an object that is referred to by a spoken word captures attention. To test this prediction, we examined whether a task-irrelevant spoken word can disrupt performance in a visual detection task requiring an eye-movement response (experiments 1 and 2) or in a visual discrimination task requiring a manual response (experiment 3). As outlined in the introduction, attentional capture is the rapid and reflexive orientation of attention toward a stimulus. We subscribe to an operational definition of attentional capture put forward by Yantis (1993, p. 679): "Stimulus-driven attentional capture can be said to occur only when the attribute that elicits it is independent of the defining and reported attributes of the target." In each of our experiments, we examined whether performance on a simple visual task involving a display with several objects would be affected by the presentation of a spoken word that referred to an object in the visual display and whose onset immediately preceded (by 400 ms) the task-relevant visual stimulus.

In each of the three experiments, participants were required to process basic visual properties of objects: Participants were required to respond to a color change (in experiments 1 and 2) or the onset of motion (in experiment 3). Importantly, the task-relevant visual cue that the participant responded to was independent of any information that might be activated upon hearing a concurrently presented spoken word. Moreover, a response to a change in color or the onset of motion does not require accessing higher-level information, such as an object's identity or other semantic information that might be activated upon seeing an object and/or hearing its name. Studies using visual matching tasks have shown that when an observer attends to an object's basic visual features, such as color or motion, this does not (automatically) result in semantic processing of the object (Boucart, Humphreys, & Lorenceau, 1995; Humphreys & Boucart, 1997). There is evidence that listeners activate visually based conceptual information upon hearing a spoken word, and that this information is used to map speech onto potential referents in visual-world studies (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007). Importantly, however, in the current set of experiments such information could interfere with the participant's task (i.e., color detection or motion discrimination) only through influencing the allocation of visual attention. This is because the defining attribute of the target object (color in experiments 1 and 2, and motion in experiment 3) is independent of any attributes of the target object that might be cued by the spoken word.

Experiment 1

In experiment 1, participants generated an eye movement in response to an isoluminant color change. They were presented with a visual display including two objects and a central fixation cross. After a short delay, a spoken word was presented, which referred to the target object (*congruent* condition) or to the other object (*incongruent* condition). Four hundred ms after the onset of the word, one of the objects turned green. Participants were instructed to move their eyes to the target object as soon as they detected the color change. Of interest was whether this overt shift in visual attention would be affected by whether or not the spoken word referred to the target object (note that participants were told explicitly to ignore the spoken word). Before the execution of an eye movement, attention is covertly shifted to the area of interest (Shepherd, Findlay, & Hockey, 1986). If an object that is referred to by a spoken word captures attention, we would therefore expect that visual attention would shift toward that object, even in the absence of an overt eye movement. In the incongruent condition, when the word refers to the distractor object, this covert shift in attention should interfere with the shift in attention that is required for the initiation of an eye movement toward the target, because the latter process would require a reorientation of visual attention. On the assumption that reorienting attention is associated with a cost, we expected that involuntarily shifting attention toward the referent of a spoken word would affect the time it takes to initiate the eye movement to the target object cued by the color change, and would do so even when an eye movement had not been initiated toward the named object prior to the color change. Thus, if an eye movement to the target object is affected by a concurrently presented spoken word that refers to an object in the visual field, this would suggest that listeners automatically shift their attention to the visual referent of the spoken word.

Participants

Sixteen students at the University of Rochester, all native speakers of American English, took part in the experiment. They reported normal or corrected-to-normal vision and normal hearing.

Stimuli

Eye movements were recorded from the participant's right eye using a head-mounted SR Research EyeLink II eye-tracking system, sampling at 250 Hz. Stimuli were presented on an NEC MultiSync FP 1305X color monitor. Stimulus presentation was controlled by a PC running ExBuilder experimentation software (Longhurst, 2006). Average stimulus luminance was measured using a Minolta LS-110 photometer. Spoken stimuli were presented binaurally through Sennheiser HD 570 headphones.

Stimuli consisted of 60 pairs of objects, selected from a standardized set of line drawings of familiar objects (Snodgrass & Vanderwart, 1980). For the purposes of the experiment, we created a monochrome gray and a monochrome green version of each object. To ensure that the gray to green color change (see procedure below) was not associated with a change in luminance, both

colors were calibrated at 15.0 cd/m^2. This was done because changes in luminance are salient cues which have the potential to capture attention (Irwin, Colcombe, Kramer, & Hahn, 2000). Images were presented on a white background (mean luminance of 38.0 cd/m^2). Within each stimulus pair, object names were phonologically dissimilar (e.g., *cat* and *sun*). The names of each of the $2 \times 60 = 120$ objects were recorded in randomized order by a male native speaker of standard American English. Words ranged in duration from 372 to 830 ms, with an average word duration of 593 ms ($SD = 93$ ms).

Design

The experiment consisted of 12 practice trials followed by 48 experimental trials. Four lists were created by varying, for each of the 48 experimental trials, which object changed color and which object was referred to by the spoken word. Thus, in one half of the trials, the spoken word referred to the target object (*congruent* condition), whereas in the other half of the trials, it referred to the distractor object (*incongruent* condition). Similarly, in one half of the trials, the spoken word referred to the object on the left side of the fixation cross, whereas in the other half of the trials, it referred to the object on the right side of the fixation cross. Four randomizations were created by varying the order of experimental trials. Each of these randomizations was applied to each of the four lists, yielding a total of 16 different randomized trial lists. Each participant was assigned to one trial list.

Task and Procedure

Participants were seated at approximately 50 cm from the computer screen. The lab was dimly lit; the only two sources of light were the monitor used for stimulus presentation and the monitor attached to a PC running the eye-tracker, which was located behind the participant. At the start of the experiment, the eye-tracker was fitted and calibrated. Participants received instructions that they would see two objects on the computer screen, and that one object would turn green. They were instructed to fixate a central cross at the beginning of each trial, and to move their eyes as quickly as they could from the central fixation cross to the object that turned green. Participants received explicit instructions to "make an eye movement in response to the color change, not in response to the words that are being spoken."

The structure of a trial was as follows. First, a fixation cross ($0.5° \times 0.5°$) appeared in the center of the screen. The participant carefully fixated the cross and pressed the spacebar to trigger a drift correction of the eye-tracking system followed by the initiation of the trial. After a delay of a second, the fixation cross turned into an asterisk ($0.5° \times 0.5°$) and two gray objects appeared (Fig. 1). Each object subtended approximately $8° \times 8°$. The objects were centered approximately 8° from fixation; the distance between the centers of the objects was thus approximately 16°. After a delay of 3 seconds, a spoken object name was presented which referred to one of the two objects on the computer screen. Four hundred milliseconds after the onset of the spoken word, the asterisk turned into a fixation cross and one of the two objects turned green. The objects remained on the screen for 2 seconds following the color change. Subsequently, a blank screen was displayed for 2 seconds before the beginning of the next trial.

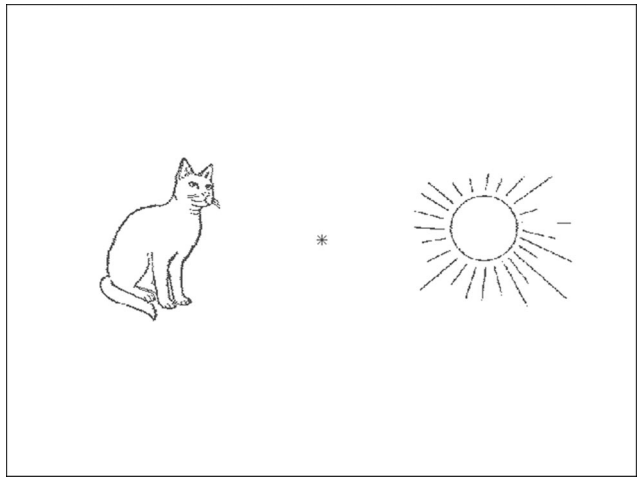

Figure 1. Example of a visual display used in experiment 1. The spoken distractor word was "cat" or "sun."

Results

Twelve trials (1.6% of the data) were discarded because the first saccade following the color change did not land within an area of $8° \times 8°$ from the center of one of the objects. Thirty additional trials (3.9% of the data) were discarded because the participant generated a saccade to the distractor object instead of the target object. We then computed, for each experimental trial, the saccade latency, that is, the time it took the participant to initiate an eye movement to the target object following the color change. Trials with unusually short or long saccade latencies were excluded from further analysis: 14 trials with a latency of less than 150 ms (1.8% of the data) and 13 trials with a latency of more than 500 ms (1.7% of the data). These outlier criteria were chosen on the basis of a visual inspection of a histogram of saccade latencies. Note that including the outliers in the analysis did not change the statistical patterns reported below.

We analyzed the remaining trials, on which participants made a correct response by generating a fixation to the target object (91.0% of the data). Of interest was whether there would be a difference in saccade latencies to the target object between the *congruent* condition, where the spoken word referred to the target object, and the *incongruent* condition, where the spoken word referred to the distractor object. On average, latencies were 16 ms slower in the *incongruent* condition (308 ms) than in the *congruent* condition (292 ms). Saccade latencies were log transformed and subjected to a multilevel regression model with random intercepts for participants and items, and condition as fixed effect. The effect of condition was significant[2] ($\beta = .047$, $p_{\mathrm{MCMC}} < .0005$). Throughout this paper, multilevel regression models were fit using the *lmer()* function of the *lme4* package (Bates & Maechler, 2010) in the free statistics software R (version 2.10.1; R Development Core Team, 2010). *p*-values for fixed effects were estimated and reported as the posterior probability of a Markov Chain Monte

[2] The effect of condition was also significant ($p < .05$ or smaller) when the data were analyzed with a two-tailed paired *t* test on log-transformed saccade latencies averaged across participants or items.

Carlo (MCMC) simulation with 10,000 samples (see Baayen, Davidson, & Bates, 2008, for discussion). The regression models contained the maximum random-effect structure justified by the data, as assessed by comparing a model with random intercepts for participants and items with a model with a random intercept for participants only, and a model with a random intercept for items only.

In an additional and more conservative analysis of the data, we only considered trials on which the participant did not move their gaze from the fixation cross prior to the color change in response to hearing the word, and was fixating the cross at the moment of color change. This analysis thus included only those trials on which there was no evidence for an overt shift in attention toward one of the objects upon hearing the spoken word. (For this purpose, and throughout the paper, fixations falling within an area of 1.5° from the center of the fixation cross were considered fixations to the cross.) Seventeen trials on which the participant moved their eyes away from the cross after the onset of the spoken word but prior to the color change were excluded (2.2% of the data; this included four trials on which the participant fixated one of the objects). Twenty-two trials on which the participant was not fixating the cross at the time of the color change were also discarded (2.9% of the data).

The remaining trials corresponded to 85.9% of the data. Average saccade latencies were 16 ms slower in the *incongruent* condition (307 ms) than in the *congruent* condition (291 ms). Saccade latencies were log transformed and subjected to a multilevel regression model with random intercepts for participants and items, and condition as fixed effect. The effect of condition was significant ($\beta = .050$, $p_{MCMC} < .0005$). Taken together, these results demonstrate that eye movements to the target object were initiated more slowly when the spoken word referred to the distractor object than when the spoken word referred to the target object. This suggests that the object that was referred to by the spoken word captured attention.

Discussion

In experiment 1, participants performed a simple and purely visual task where the only requirement was to make an eye movement toward an object that changed color. Participants had been explicitly instructed to ignore spoken distractor words, whose onsets preceded the color change by 400 ms. Nevertheless, our results demonstrate that the spoken words systematically affected performance on the visual task. Eye movements to the target object were initiated faster when the spoken word referred to the target object compared to when the spoken word referred to the distractor object. This result suggests that visual attention was allocated to the objects in the visual environment that were referred to by the spoken words. The next two studies examine in more detail whether we can deem such attention allocation as "automatic."

In experiment 1, we found a linguistic interference effect even though the spoken distractor words were irrelevant to the experimental task. The distractor word was equally likely to refer to the target object as to the distractor object and therefore did not constitute a valid cue to the location of the target object. However, even though the spoken word was not a valid cue to the location of the color change (i.e., it was a valid cue on half of the trials and an invalid cue on the other half of the trials), it is conceivable that (some) participants may have treated the spoken words as cues to try to improve their performance by anticipating the location of the color change. Anecdotal reports from participants suggest that this is not the case: many participants reported that the spoken words were very distracting, and that they had trouble trying to ignore the words. However, the fact that the spoken word frequently referred to the object that changed color may have encouraged, or contributed to, the observed interference effect. Thus, from (some of) the participants' perspective, the language may not in fact have been task-irrelevant—the "strategic" use of a cue in service of a task or goal, even if unintended by the experimenter, would compromise the interpretation of the modulating effect of that cue as "automatic," that is, task-independent.

In experiment 2, we addressed concerns about the potential use of strategies by participants in our initial study. The task in experiment 2 was identical to that in experiment 1: to generate, as quickly as possible, an eye movement to an object in a visual display that turned green. We made two modifications to the design of our initial experiment. First, visual displays included four objects. Second, the spoken distractor word never referred to the target object. The distractor word either referred to an object in the visual display that did not change color (*present* condition) or to an object that was not included in the visual display (*control* condition). We predicted that the presentation of a distractor word that referred to an object in the display would delay the time it took participants to generate an eye movement to the object that changed color. If response times are slower in the *present* condition than in the *control* condition this would suggest that in the *present condition*, participants directed their attention to the visual referent of the spoken word and that this shift in attention interfered with the initiation of an eye movement to the target object in response to the color change. Crucially, the spoken words in this study had zero validity in respect of predicting the task-relevant target.

Experiment 2

Participants

Sixteen students at the University of Rochester took part in the experiment. They were all native speakers of American English. Participants reported normal or corrected-to-normal vision and normal hearing.

Apparatus and Stimuli

The apparatus was the same as that used in experiment 1. Stimuli consisted of 60 sets of 4 objects, selected from a standardized set of line drawings of familiar objects (Snodgrass & Vanderwart, 1980). One object was designated the role of target object, that is, the object that would change color. The three distractor objects were chosen carefully such that their names were phonologically dissimilar to the name of the target object. A control word was selected for each set of objects. The control word was phonologically dissimilar to the names of the objects and closely matched to the target word in lexical frequency and word length. It was important to match target and control words so that, for each item set, potential differences in responses between the *present* and *control* condition could not be explained by differences in lexical

properties of the words associated with each condition. The average frequency of the target word (estimated from frequency counts reported in Francis and Kučera, 1982) was 21.1 per million, and the average frequency of the control word was 21.0 per million. A paired t test indicated that the difference in log-transformed frequency between target words and control words (2.49 vs. 2.52) was not significant, $t < 1$. The average length of the target word was 4.3 phonemes, and the average length of the control word was 4.5 phonemes. This difference in word length was not significant in a paired t test, $t = 1.6$, $p > .1$.

The $60 \times 2 = 120$ words were recorded in randomized order by a male native speaker of American English. The duration of each word was measured using speech editing software. Average word duration was 579 ms (SD = 79 ms) for words in the *present* condition and 577 ms (SD = 93 ms) for words in the *control* condition. A paired t test indicated that this difference was not significant ($t < 1$).

Design

The experiment consisted of 12 practice trials followed by 48 experimental trials. Two lists were created by varying which spoken word was presented along with a visual display. In one half of the trials, the word referred to an object in the display (*present* condition), whereas in the other half of the trials, the spoken word did not refer to an object in the display (*control* condition). For each set of items, picture positions were determined quasi-randomly so that throughout the experiment there were four instances of each of the 12 possible combinations of target position and named distractor position. Four randomizations were created by varying the order of experimental trials. Each of these randomizations was applied to each of the two lists, yielding a total of eight randomized lists. Two participants were randomly assigned to each randomized list.

Task and Procedure

The procedure was identical to experiment 1, except that the visual display included four objects, with one object appearing in each quadrant of the screen. Each object was approximately 6° × 6° in size. The distance between the central fixation cross and the center of each object was approximately 8.5°. The distance between the centers of two adjacent objects was approximately 10.5°. Participants were instructed to fixate the central cross at the beginning of each trial, and to move their eyes, as quickly as they could, to the object that turned green. They received explicit instructions to "make an eye movement in response to the color change, not in response to the words that are being spoken."

Results

Trials on which the first saccade following the color change did not land within an area of 6° × 6° from the center of an object were discarded (133 trials; 17.3% of the data). Forty-seven additional trials (6.1% of the data) were discarded because the participant fixated the distractor object following the color change. For the remaining trials, we computed the saccade latency, that is, the time it took the participant to initiate an eye movement to the target object following the color change. Trials with unusually short or long saccade latencies were excluded from further analysis: two trials with a latency of less than 150 ms (0.3% of the data) and eight trials with a latency of more than 700 ms (1.0% of the data). These outlier criteria were chosen on the basis of a visual inspection of a histogram of saccade latencies. Including these outliers in the analysis did not change the statistical patterns reported below.

We analyzed the remaining trials, on which participants made a correct response by generating a fixation to the target object (75.3% of the data). Of interest was whether there would be a difference in saccade latencies to the target object between the *present* condition, where the spoken word referred to a distractor object, and the *control* condition, where the spoken word referred to an object not in the display. On average, saccade latencies were 22 ms slower in the *present* condition (358 ms) than in the *control* condition (336 ms). Saccade latencies were log transformed and subjected to a multilevel regression model with random intercepts for participants and items, and condition as fixed effect. The effect of condition was significant (see footnote 2) ($\beta = .067$, $p_{\mathrm{MCMC}} < .0005$).

In an additional and more conservative analysis of the data, we only considered trials on which the participant did not move their gaze from the fixation cross prior to the color change in response to hearing the word, and was fixating the cross at the moment of color change. Eleven trials on which the participant moved their eyes away from the cross after the onset of the spoken word but prior to the color change were excluded (1.4% of the data, including one trial on which the participant fixated one of the objects). Twenty-one trials on which the participant did not fixate the cross at the time of the color change were also discarded (2.7% of the data).

The remaining trials corresponded to 71.1% of the data. Average saccade latencies were 20 ms slower in the *present* condition (358 ms) than in the *control* condition (338 ms). Saccade latencies were log transformed and subjected to a multilevel regression model with random intercepts for participants and items, and condition as fixed effect. The effect of condition was significant ($\beta = .062$, $p_{\mathrm{MCMC}} < .0005$). Taken together, these results demonstrate that eye movements to the target object were initiated more slowly when the spoken word referred to a distractor object than when the spoken word referred to an object not in the display.

Discussion

In experiment 2, participants saw an array of four objects and again performed a simple and purely visual task: to generate an eye movement toward whichever object turned green. Just like in experiment 1, participants received explicit instructions to ignore the spoken distractor word, whose onset preceded the color change by 400 ms. The distractor word never referred to the object that changed color and was thus not a valid cue to the location of the target object on any trial. Nevertheless, we found that the spoken distractor word interfered with the visual task when it referred to one of the objects in the visual display: Reaction times for correct responses were 20 ms slower when the spoken word referred to a distractor object in the display than when it did not. (The design of this study does not permit us to conclude that when the word referred to some other object, not depicted in the array, there was no distraction effect.) This finding suggests that a spoken word that refers to a distractor object results in an involuntary shift of

attention toward this object, which in turn slows down the initiation of an eye movement to the target object. The results of experiment 2 thus converge with those of experiment 1 in suggesting that objects referred to by spoken language capture attention.

There is one possible objection to our assessment of the spoken word as not constituting a valid cue to the location of the color change. On half the trials, that is, in the *present* condition, the spoken word always referred to one of the distractor objects. Therefore, on those trials, the object that the spoken word referred to would *not* change color. By identifying that object, participants could in principle subsequently narrow down the set of objects that needed to be considered in order to detect the color change. However, if participants were able to use this strategy to improve their performance, we would expect to have obtained faster response times in the *present* condition than in the *control* condition—contrary to what was observed. For this reason, and in the context of explicit instructions to generate a response as quickly as possible, we believe that it is unlikely that the effect of condition observed in experiment 2 reflects intentional, rapid, and successful exclusion of the named distractor from the response set.

There are at least two important limitations of the studies presented thus far. A first limitation is that the effect of spoken object names on the allocation of visual attention is measured indirectly. The results show that the auditory presentation of the name of an object in the visual display affects the time it takes to initiate a saccade to a target object. The assumption is that an eye movement is preceded by a shift in visual attention to the location that will subsequently be fixated. If attention automatically shifts to the referent of a spoken word, this should make it easier to generate an eye movement to that object if that object subsequently changes color, while it should make it harder to generate an eye movement to a different object if that different object subsequently changes color. The idea is that generating an eye movement to the object referred to by the spoken word is easy due to the fact that attention is already oriented on that object, while generating an eye movement to an object other than the one referred to by the spoken word is harder because it requires a reorientation of visual attention. Nevertheless, the experimental task itself measures the allocation of visual attention only indirectly, that is, through the effect that orientation of visual attention has on the subsequent initiation of an eye movement.

It is well known that the primary effect of allocating attention to a specific location in the visual field is that it enhances processing of information at that location. Thus, a more direct test of the automaticity of language-mediated attention shifts would be to find evidence for attentional capture of objects referred to by spoken language in a task that directly measures the ease of processing of visual information at a particular location. In experiment 3, participants performed a visual-discrimination task which required a perceptual judgment resulting in a manual response, instead of an eye-movement response. The goal of this experiment was to examine directly if hearing the name of an object affects the processing of visual information associated with that, and other, objects.

A second limitation of experiments 1 and 2 is that the results are somewhat inconclusive due to the experimental design. Results from both experiments strongly suggest that visual attention is allocated to an object when the name of that object is heard. But evidence concerning the mechanism underlying this effect is somewhat equivocal. In experiment 1, participants were faster to initiate an eye movement to the target object when the spoken word referred to the target object than when the word referred to the distractor object. This difference in saccade latencies between the *congruent* and *incongruent* conditions is consistent with a facilitatory effect (when the word refers to the target object), an inhibitory effect (when the word refers to the distractor object) or both. In experiment 2, participants were slower to initiate an eye movement to the target object when the spoken word referred to one of three other objects in the visual display than when the word referred to an object not in the display. This difference in response times between the *present* and *control* conditions is consistent with an inhibitory effect (when the word referred to an object in the display).

Taken together, the results of experiments 1 and 2 are consistent with the idea that hearing the name of a distractor object inhibits, or interferes with, processing of a target object and/or programming of an eye movement to that object. However, the results do not provide conclusive evidence that hearing the name of a target object *facilitates* processing of that object. In experiment 3, we examined the nature of the effects of the task-irrelevant language in more detail. For this purpose, we introduced a baseline condition in order to assess the nature of the interference effect due to task-irrelevant distractor words in a visual task. In particular, the baseline condition allowed us to assess whether hearing the name of an object can facilitate as well as inhibit the processing of visual information associated with the target object.

Experiment 3 employed a visual-discrimination task. Participants saw two objects on the computer screen and fixated a central fixation cross. After a delay of 3 seconds, a spoken word was presented. Four hundred ms after the onset of the spoken word, one of the two objects shifted slightly up or down. The task of the participant was to determine, as quickly and accurately as they could, the direction of shift (up or down). The visual display disappeared approximately 60 ms after the target object had shifted. This was done to prevent participants from making an eye movement to the object that shifted to assist their response. Spoken distractor words were presented to assess the influence of spoken object names on the allocation of visual attention. The distractor word matched the object that shifted (*congruent* condition), the object that did not shift (*incongruent* condition), or an object not on the screen (*control* condition). Importantly, the *control* condition provides a baseline that allows to establish whether the effects we have observed previously are likely due to inhibitory effects, facilitatory effects, or both.

Experiment 3

Participants

Eighteen students from the University of Rochester took part in the experiment. They were all native speakers of American English and reported normal or corrected-to-normal vision.

Apparatus and Stimuli

The apparatus was the same as that used in experiments 1 and 2. Stimuli consisted of 60 visual displays, corresponding to 12 prac-

tice displays and 48 experimental displays. Each display consisted of two objects, selected from a standardized set of black-and-white line drawings of familiar objects (Snodgrass & Vanderwart, 1980). Three words were associated with each of the 48 experimental displays: the names of the two objects, and a control word. Within each visual display, the two object names were phonologically dissimilar (e.g., *dog*, *gun*). The control word was phonologically dissimilar from each of the two object names (e.g., *bed*) and corresponded to the name of an object that was not present in the visual display. For each visual display, the names of the two objects and the control word were chosen so that they were closely matched in frequency and word length. Word frequencies were estimated from frequency counts reported in Francis and Kučera (1982). Average frequencies were 26.1 per million for the name associated with the left object, 26.1 per million for the name associated with the right object, and 26.0 for the control word. An ANOVA showed that differences in log-transformed frequency between the three groups of words were not significant ($F(2, 94) < 1$). The average length of the name associated with the left object was 4.0 phonemes, that of the name associated with the right object 4.1 phonemes, and that of the control word 4.2 phonemes. An ANOVA showed that differences in word length between these three groups of words were not significant, $F(2, 94) = 1.0$, $p = .368$.

One word was associated with each of the 12 practice displays. This word referred to the left object for four displays, the right object for four displays, and to an object not in the display for four displays. The names of each of the $(3 \times 48) + 12 = 156$ objects were recorded in randomized order by a male native speaker of standard American English. Words recorded for experimental trials ranged in duration from 400 to 875 ms, with an average duration of 558 ms ($SD = 84$ ms). The average duration of the name of the left object was 557 ms, that of the right object 551 ms, and that of the control object 573 ms. An ANOVA showed that differences in duration between these three groups of words were not significant ($F(2, 94) < 1$).

Design

Twelve lists were created by varying, for each of the 48 experimental displays, which of the two objects shifted, whether the object shifted up or down, and which object the spoken word referred to (i.e., the target object, the distractor object, or an object not on the screen). Each list had one of the 12 possible combinations of these factors associated with a particular experimental display. Within each list, four experimental displays were associated with one of the 12 possible combinations of target object (left or right), direction of shift (up or down) and condition (congruent, incongruent, or control). The order of trials was randomized for each of the 12 lists, with the constraint that the direction of shift of the target object was not identical on more than four consecutive trials. A fixed set of 12 randomly ordered practice trials preceded the 48 experimental trials, with one practice trial associated with each of the 12 possible combinations of factors detailed above. One or two participants were assigned to each randomized list.

Procedure

Participants were seated at approximately 50 cm from the computer screen. At the start of the experiment, the eye-tracker was fitted and calibrated. Participants were instructed to fixate a central cross throughout each trial. Their task was to identify the direction of movement of the object that shifted by pressing a key on the keyboard (the keys "Y" and "H" in the center of the keyboard were labeled with an arrow pointing upwards and downwards, respectively). Participants were asked to keep the index fingers of their left and right hand on the response keys throughout a trial. They were instructed to respond as quickly and accurately as they could, without moving their eyes from the fixation cross. Participants were informed that they would hear spoken words and were instructed explicitly that "these words are not relevant to the task that you are performing."

The structure of a trial was as follows. First, a fixation cross ($0.5° \times 0.5°$) appeared in the center of the screen, flanked by two objects. Each object subtended approximately $7° \times 7°$, and the objects were centered approximately 6° from fixation. The distance between the centers of the objects was approximately 12°. After a delay of 3 seconds, a spoken word was presented through headphones. The word referred to the target object (*congruent* condition), the distractor object (*incongruent* condition) or to an object not on the screen (*control* condition). Four hundred ms after the onset of the spoken word, the target object shifted slightly up or down by about 0.5° (12 pixels, using a screen resolution of $1{,}024 \times 768$ pixels). Five screen refreshes later (i.e., after ~59 ms, given a monitor refresh rate of 85 Hz) the screen went blank. Once the participant indicated their response by pressing a key on the keyboard, there was a delay of 2 seconds before the experiment advanced to the next trial. Every four trials, a small dot appeared in the center of the screen. The participant carefully fixated the dot before pressing the space bar to trigger a drift correction for the eye-tracking system.

Results

Due to slight variation in the timing of stimulus presentation, which was beyond our control, there were 63 trials (7.3% of the data) on which the objects remained on the screen for six instead of five screen refreshes after the target object had shifted. These trials were included in the analyses. (Note that the same statistical patterns as described below were found when these trials were excluded from the analyses.) Seven trials (0.8% of the data) were discarded because the participant pressed the space bar or another key on the keyboard instead of one of the two response keys. Trials with a response time of more than two standard deviations above the mean response time were considered outliers. These trials were excluded from further analysis (23 trials, including nine trials with an incorrect response; 2.7% of the data).

Correct Responses

One hundred forty-two trials (16.4% of the data) were discarded because the participant's response did not match the direction in which the target object shifted. (Given the large number of incorrect responses, we analyzed these trials separately; see below). Trials on which participants correctly indicated the direction in which the target object shifted corresponded to 80.1% of the data. Average response times were computed for each of the experimental conditions. Responses were fastest in the *congruent* condition (707 ms), intermediate in the *control* condition (738 ms),

and slowest in the *incongruent* condition (817 ms). Response times were log-transformed and subjected to a multilevel regression model with random intercepts for participants, and condition as fixed effect (using treatment coding, with the control condition as the baseline). Nested model comparison indicated that including random intercepts for items did not significantly improve the model. This suggests that there was little variation in the time it took participants to make a response as a function of which particular object (out of the 96 objects comprising experimental trials) shifted. The effect of condition was significant when comparing the *congruent* condition to the *control* condition ($\beta = -.067$, $p_{MCMC} < .05$), as well as when comparing the *incongruent* condition to the *control* condition ($\beta = .088$, $p_{MCMC} < .01$).[3]

In an additional and more conservative analysis of the data, we only considered trials on which the participant did not move their gaze from the fixation cross prior to the shift in object position in response to hearing the word, and was fixating the cross when the target object shifted. Forty trials on which the participant moved their eyes away from the cross after the onset of the spoken word but prior to the shift were excluded (4.6% of the data, including 18 trials on which the participant fixated one of the objects). Forty-eight trials on which the participant did not fixate the cross at the time of the shift in object position were also discarded (5.6% of the data).

The remaining trials corresponded to 69.9% of the data. Responses were fastest in the *congruent* condition (698 ms), slower in the *control* condition (734 ms), and slowest in the *incongruent* condition (825 ms). Response times were log-transformed and subjected to a multilevel regression model with random intercepts for participants and condition as fixed effect (using treatment coding, with the control condition as the baseline). The effect of condition was significant when comparing the *congruent* condition to the *control* condition ($\beta = -.064$, $p_{MCMC} < .05$), as well as when comparing the *incongruent* condition to the *control* condition ($\beta = .099$, $p_{MCMC} < .005$).

Incorrect Responses

To assess if the linguistic distractors influenced accuracy, we examined those trials on which participants incorrectly indicated the direction in which the target object moved. In accordance with the "more conservative" analysis of correct trials reported above, we only considered trials on which the participant did not move their gaze from the fixation cross prior to the shift in object position in response to hearing the word (resulting in the exclusion of 21 trials), and was fixating the cross when the target object shifted (resulting in the exclusion of 11 trials). This was done to ensure that an incorrect response was not due to an overt shift of attention to (or toward) the target or distractor object, or to the participant not fixating the cross at the moment when the target object shifted.

The proportion of incorrect responses was 16.0% in the *control* condition (38 out of 238 responses). Compared with the control condition, performance was more accurate in the *congruent* condition (8.7% incorrect; 20 of 229 responses), and less accurate in the *incongruent* condition (21.1% incorrect; 52 of 247 responses). Incorrect responses were subjected to a multilevel logit model, which predicted the log odds of observing a correct response. The model included random intercepts for participants, and condition as fixed effect (using treatment coding, with the control condition as the baseline). The effect of condition was significant when comparing the *congruent* condition to the *control* condition ($\beta = .68$, $p < .05$), and marginally significant when comparing the *incongruent* condition to the *control* condition ($\beta = -.42$, $p = .087$). (Note that these p values are based on two-tailed z tests). These results suggest that on those trials where there was no evidence that participants shifted their gaze to or toward either of the objects upon hearing the distractor word, participants determined the direction of shift of the target object more accurately when the spoken distractor word referred to that object. Thus, the results of this analysis converge with the response time analysis: Determining the direction of motion of the object that shifted was *facilitated* when the spoken word referred to that object, and *inhibited* when the spoken word referred to the other object. Taken together, these analyses suggest that when a spoken word referred to an object in the display, attention was rapidly and automatically allocated toward that object.

General Discussion

Our results show that visual attention is influenced by linguistic processing. Performance on a simple visual detection or discrimination task was affected by the presentation of the spoken name of an object in the visual field. In two visual-detection experiments, participants saw a visual display and generated an eye movement to an object that turned green. In experiment 1, the visual display included two objects, one of which turned green. The onset of a spoken distractor word which referred to one of the objects preceded the color change by 400 ms. Participants were slower to initiate an eye movement to the target object when the spoken word referred to the distractor object than when the spoken word referred to the target object. In experiment 2, the visual display included four objects. A spoken distractor word referred to one of the distractor objects (i.e., an object that did not change color), or to an object not included in the visual display. Although the spoken word never referred to the target object, participants were slower to initiate an eye movement to the target object when the spoken word referred to one of the distractor objects than when the spoken word did not have a referent. Experiment 3 replicated the linguistic interference effect observed in experiments 1 and 2 with a discrimination task that required a manual response. Participants saw two objects in a visual display and fixated a central cross. One of the objects shifted slightly up or down, and participants had to indicate the direction of shift. A spoken distractor word referred to one of the objects or to an absent object. Participants were fastest to detect the direction of shift of the target object when the spoken word referred to that object, slower when the spoken word did not have a referent, and slowest when the spoken word referred to the distractor object. Taken together, the results from these three experiments converge in showing that task-irrelevant spoken object names can interfere with performance on a simple visual task.

Importantly, we found that spoken distractor words affected performance in a visual task even though the one and only task of

[3] These effects were also significant ($p < .05$) when the data were analyzed with a two-tailed paired t test on log-transformed response times averaged across participants.

the participants was to respond to the visual cue, and the experimental instructions explicitly stated that the spoken words should be ignored. Our results thus provide stronger evidence for automatic influences of linguistic information on visual attention than prior experiments where participants were required to process linguistic information as part of the experimental task (Soto & Humphreys, 2007). In our experiments, there was no incentive for participants to process the spoken words, and many participants reported that they actively tried to suppress processing of the spoken words once they noticed that the words interfered with the visual task that they were trying to perform. Nevertheless, our results clearly show that the spoken words interfered systematically with the experimental task, even though these words were task-irrelevant, suggesting that objects that are referred to by spoken words capture attention. Although we did not manipulate cue validity systematically (to show language-modulation of visual attention independently of cue validity), cue validities across the three experiments were 50% (experiment 1), zero (experiment 2), or 33% (experiment 3). Moreover, the object referred to by the spoken word was no more likely to be the target than the other object(s) in the display. Yet, despite variation in cue validity, the data consistently showed language mediation of attentional allocation, which is consistent with the idea that the object referred to by the spoken word captured attention (see Ellsiepen, Ferreira, & Henderson, submitted, for related and converging evidence, in a study that focused on manipulating cue validity of spoken distractors). We conclude, therefore, that language-mediation of the allocation of visual attention is automatic, *defined operationally*—in terms of its mediation irrespective of task-relevance.

One could argue that the linguistic interference effect that we observed is obtained only because participants deliberately shifted their attention to whichever object in the visual display matched the spoken distractor word, in anticipation of the visual cue. This account of our results would assume that participants, for some reason, adopted a strategy that violated the instructions provided at the beginning of the experiment. A priori, it is not clear why participants would decide on such a strategy, which is particularly hard to motivate in the context of experiment 2, where the spoken distractor word never referred to the target object. Moreover, a large body of research on attentional capture of visual cues has shown that observers readily adopt specific attentional control settings that are appropriate for the task that they are required to perform. For instance, a classic study by Folk, Remington, and Johnston (1992) showed that abrupt onsets do not capture attention when the search target is defined as a color singleton. Similar results have been obtained in many related studies, suggesting that stimuli that do not match the attentional control setting required for the experimental task, and which are therefore irrelevant to the observer's goals, do not capture attention. In our experiments, participants were required to adopt an attentional control setting for the color green (in experiments 1 and 2) or movement (in experiment 3). These basic visual features are easy to detect, as properties of objects, and typically do not result in semantic processing of those objects (Boucart et al., 1995; Humphreys & Boucart, 1997). In light of results from the attentional control setting literature showing that participants can readily ignore visual stimuli, including abrupt onsets that otherwise (for instance, when no clear attentional control setting is provided by the task) capture attention, there is no clear explanation why participants in our studies might have decided to voluntarily shift their attention to the object that the spoken word referred to—counter to the experimental instructions. Taken together, our results support the hypothesis that objects, when referred to by a spoken word, capture attention.

The classic version of the Stroop task (see MacLeod, 1991, for a review) famously demonstrates that when a printed word is perceived, its meaning is accessed rapidly and automatically. In the standard version of this task, participants are asked to report the color of a series of printed words. Invariably, the time it takes a participant to produce the relevant color name is affected by the meaning of the word. For instance, participants are faster to report that a word is printed in green when the word is *sock* than when the word is *blue*. Our results bear a resemblance to the Stroop task in showing that lexical semantics (the meaning of a word) can interfere with the execution of a task in which this knowledge is not relevant. However, our results go significantly beyond demonstrating a linguistic Stroop effect in the auditory domain.

Our findings constitute more than a demonstration that spoken words, like their written counterparts, are processed and recognized automatically. The interference effect that we observed is mediated by referential processing of spoken language in a visual context. Reference resolution is a core aspect of language processing, and our results suggest that this process can impinge on the allocation of attention in the visual system; whether or not the spoken word referred to an object *in the display*, and which object it referred to, determined the nature of the interference effect. Our data also go beyond the classic Stroop effect because in the Stroop task, interference effects arise through featural mismatch—for instance, the color of the letters mismatching with the color that the word refers to. In our studies, there was no such featural mismatch: the visual response (manifest as eye movements in the first two experiments and as a key press in the third experiment) was contingent on a visual feature that was orthogonal to the featural dimensions along which the spoken words could be described. The featural match or mismatch that drove our effects was between the location of the visual target and the location of the object to which the spoken word referred. Crucially, the effects we observed depended on whether the word referred to an object in a *specific* location that matched or mismatched the task-relevant location to which attention had to be directed.

Notwithstanding the differences between our effects and the interference effects observed in Stroop, there are also important similarities, at least with respect to theoretical interpretation. In Stroop interference studies, the participant must attend to one featural dimension, such as color, at the expense of another, such as the meaning of the printed word. In our studies, participants had to attend to the visual change at the expense of attending to the spoken word. Like in Stroop interference, the automatic attending to that spoken word resulted in interference (experiments 2 and 3, and possibly experiment 1) or facilitation (experiment 3, and possibly experiment 1). In other words, like in Stroop interference studies, participants in our studies had to attend to the visual features in their environment, but did not do so at the *total* expense of attending to the spoken words. Stroop interference has been interpreted theoretically in the context of the *guided activation theory of cognitive control* (Miller & Cohen, 2001; Cohen, Aston-Jones, & Gilzenrat, 2004) and the *biased competition theory* (Desimone & Duncan, 1995). According to these accounts, attention can

be considered as the modulatory influence that biases the outcome of the competition between concurrently active representations. Thus, the interactions we have found between language comprehension and visual attention can be viewed as reflecting competition between the linguistic and visual representations activated during each trial. More specifically, the match and mismatch effects we observed reflected the allocation of attention in the direction of the visual change balanced against (i.e., in competition with) the language-mediated allocation of attention either in the direction of that change or in some other direction. In termed by Cohen et al (2004), "gated activation" is the influence of the context on the activation of one set of featural representations or another (they can be overlapping)—where, "context" can refer to other inputs to the system (i.e. other than those causing the activation being modulated), the system's prior states, and/or the task goals. In these terms, language can act as a "gate" on attentional deployment in the visual domain.

To conclude, our results demonstrate that linguistic processing can influence the allocation of attention in the visual system very rapidly and automatically. This result may have important implications for behavior in carefully controlled operating environments where spoken language may distract an operator from attending to information that is relevant to their task (e.g., drivers or air-traffic controllers). The precise conditions under which objects referred to by spoken language capture attention in more crowded and interactive real-life environments, remain an important topic for future research. Our results show that in a simple visual environment, an object that is referred to by spoken language captures attention and interferes with the execution of a simple and purely visual task. This rapid and involuntary integration of visual and linguistic information indicates that different domains of cognitive processing show a larger degree of interactivity than is traditionally assumed.

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. (in press). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica.* doi:10.1016/j.actpsy.2010.09.009.

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

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57,* 502–518.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111,* 55–71.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Bates, D., & Maechler, M. (2010). lme4: Linear mixed-effects models using S4 classes (R package version 0.999375–34). Retrieved from http://CRAN.R-project.org/package=lme4

Belke, E., Humphreys, G. W., Watson, D. G., Meyer, A. S., & Telling, A. L. (2008). Top-down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. *Perception & Psychophysics, 70,* 1444–1458.

Boucart, M., Humphreys, G. W., & Lorenceau, J. (1995). Automatic access to object identity: Attention to global information, not to particular physical dimensions, is important. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 584–601.

Cohen, J. D., Aston-Jones, G., & Gilzenrat, M. S. (2004). A systems-level perspective on attention and cognitive control: Guided activation, adaptive gating, conflict monitoring, and exploitation versus exploration. In M. I. Posner (Ed.), *Cognitive Neuroscience of Attention* (pp. 71–90). New York, NY: Guilford 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,* 84–107.

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

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science, 11,* 467–473.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48,* 269–297.

Ellsiepen, E., Ferreira, F., & Henderson, J. M. (manuscript submitted for publication). Effects of referential words on covert visual attention.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030–1044.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston, MA: Houghton Mifflin.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.1–10.112). New York, NY: Wiley.

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., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15,* 985–1018.

Humphreys, G. W., & Boucart, M. (1997). Selection by color and form in vision. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 136–153.

Irwin, D. E., Colcombe, A. M., Kramer, A. F., & Hahn, S. (2000). Attentional and oculomotor capture by onset, luminance and color singletons. *Vision Research, 40,* 1443–1458.

James, W. (1890). *The principles of psychology* (Vol. 1). New York, NY: Henry Holt & Co.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics, 43,* 346–354.

Longhurst, E. (2006). *ExBuilder* [Computer program]. Rochester, NY: University of Rochester.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review, 14,* 710–716.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167–202.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6,* 182–189.

Moores, E., & Maxwell, J. P. (2008). The role of prior exposure in the capture of attention by items in working memory. *Visual Cognition, 16,* 675–695.

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1243–1265.

R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3–900051-07–0. Retrieved from http://www.R-project.org

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

Salverda, A. P., Brown, M., & Tanenhaus, M. K. (in press). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica.* doi:10.1016/j.actpsy.2010.09.010.

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *The Quarterly Journal of Experimental Psychology, 38,* 475–491.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 248–261.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences, 12,* 342–348.

Soto, D., & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 730–737.

Tanenhaus, M. K. (2007). Eye movements and spoken language processing. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind an brain* (pp. 443–469). Oxford: Elsevier.

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.

Yantis, S. (1993). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 676–681.

Yantis, S. (2000). Goal-directed and stimulus-driven determinants of attentional control. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII* (pp. 73–103). Cambridge, MA: MIT Press.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 601–621.

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,* 1–14.

Received March 11, 2009
Revision received December 15, 2010
Accepted December 28, 2010 ■