Perception & Psychophysics
2006, 68 (6), 959-974

Inefficient conjunction search made efficient by concurrent spoken delivery of target identity

FLORENCIA REALI, MICHAEL J. SPIVEY, MELINDA J. TYLER, and JOSEPH TERRANOVA
Cornell University, Ithaca, New York

Visual search based on a conjunction of two features typically elicits reaction times that increase linearly as a function of the number of distractors, whereas search based on a single feature is essentially unaffected by set size. These and related findings have often been interpreted as evidence of a serial search stage that follows a parallel search stage. However, a wide range of studies has been showing a form of blending of these two processes. For example, when a spoken instruction identifies the conjunction target concurrently with the visual display, the effect of set size is significantly reduced, suggesting that incremental linguistic processing of the first feature adjective and then the second feature adjective may facilitate something approximating a parallel extraction of objects during search for the target. Here, we extend these results to a variety of experimental designs. First, we replicate the result with a mixed-trials design (ruling out potential strategies associated with the blocked design of the original study). Second, in a mixed-trials experiment, the order of adjective types in the spoken query varies randomly across conditions. In a third experiment, we extend the effect to a triple-conjunction search task. A fourth (control) experiment demonstrates that these effects are not due to an efficient odd-one-out search that ignores the linguistic input. This series of experiments, along with attractor-network simulations of the phenomena, provide further evidence toward understanding linguistically mediated influences in real-time visual search processing.

For a couple of decades, the conventional wisdom on visual search was that an early, "preattentive" stage of visual processing contained multiple modules that carried out feature extraction in a parallel fashion, and that an independent attentional stage of visual processing later combined those features into object representations that were compared one at a time with a target template (Treisman & Gelade, 1980; Treisman & Sato, 1990; Woodman & Luck, 2003; see also Cavanagh, 1987; Livingstone & Hubel, 1988; Sagi & Julesz, 1984). Some of the core support for this feature integration theory has come from three rather replicable findings. First, consistent with parallel preattentive processing of individual features, search based on a single target feature (e.g., redness or verticalness) produces a perceptual pop-out, so that reaction time (RT) for target-present trials is largely unaffected by the number of distractor items in the display. Second, consistent with a serial attentional examination of each object, search based on a conjunction of two features (e.g., greenness and horizontalness) elicits RTs that increase quite linearly as a function of the number of distractor items. Third, consistent with serial self-terminating search for target-present trials and serial exhaustive search for target-absent trials when search is for a conjunction of two features, the RT slopes (msec/item) for target-present trials are often about half those for target-absent trials.

Recently, however, this stage-based serial information-processing account of visual attention has been contested by the view that visual search involves a single-stage process operating on a continuum of search efficiency. Under such an account, apparent serial search is actually the result of simultaneous competition among multiple object representations for the privilege of activating their associated motor outputs (*biased competition*; see Desimone & Duncan, 1995; Reynolds & Desimone, 2001; Spivey & Dale, 2004; Spratling & Johnson, 2004; see also Duncan & Humphreys, 1989). A special challenge for this single-mechanism framework is the apparent distinction between parallel and serial search processes. In a typical feature search, in which consultation of a single preattentive feature map is sufficient to find the target, RTs increase by less than 10 msec/item. In a typical conjunction search task, in which the target object is defined by a conjunction of features, RTs increase anywhere from 10 to 30 msec/item (Duncan & Humphreys, 1989; Treisman, 1988; Treisman & Gelade, 1980; Wolfe, 1994). This 10-msec/item threshold was often treated as an informal boundary between search functions that were referred to as "parallel" and search functions that were referred to as "serial."

It has been suggested that these different ranges of millisecond/item slopes are not the result of fundamentally different search processes (i.e., a parallel process and a

This research was supported by NIMH Grant R01-63961. We are grateful to Thomas Carr, John Henderson, Jeremy Wolfe, and an anonymous reviewer for helpful suggestions regarding the manuscript and the experiments. Correspondence concerning this article should be addressed to F. Reali or M. J. Spivey, Department of Psychology, Cornell University, Ithaca, NY 14853 (e-mail: fr34@cornell.edu or spivey@cornell.edu).

serial process) but, rather, that they emerge from a single process, perhaps determined by the relative salience of the target and the distractors (see, e.g., Dosher, Han, & Lu, 2004; Duncan & Humphreys, 1989; Eckstein, 1998; McElree & Carrasco, 1999; Palmer, Verghese, & Pavel, 2000; Wolfe, 1998). For example, Wolfe (1998) performed a meta-analysis on almost a million conjunction search and single-feature search trials and was unable to find any evidence for bimodality in the distribution of millisecond/item slopes. In a subsequent study, Haslam, Porter, and Rothschild (2001) conducted a statistical reanalysis of the same data and found that a two-distribution model of overlapping processes yielded a better fit than the one-distribution model. However, as Haslam et al. argued, their results may "complicate single mechanism or continuum models, but not refute them" (p. 745) in that a single mechanism for search could operate along an efficiency and speed continuum in which discrete changes in settling parameters could occur, or, alternatively, such a continuum could have a threshold of some sort, above which it operates in a more efficient manner. Reflecting the field's renewed uncertainty about the mechanisms underlying single-feature search and conjunction search, the visual search literature has evolved to describe these phenomena in terms of a continuum between "efficient" and "inefficient" search, in place of the dichotomy of "parallel" and "serial" search (see Nakayama & Joseph, 1998).

Some of the most compelling evidence against the independent functioning of a putative parallel "preattentive" stage of visual search and a later serial attentional stage comes from visual search displays that are temporally dynamic. Recently, a number of studies have shown that there are superficially different kinds of cues to parsing the search display and identifying the target item that have been proven to work very rapidly when delivered incrementally. Although all of these cues produce significant improvement in search efficiency, their nature may vary in that some could provide information about the location of the target, and others about its identity. For example, in one series of studies, Olds, Cowan, and Jolicœur (2000a, 2000b) presented visual search displays in the form of single-feature pop-out displays for very brief periods (e.g., 50 msec) before changing them to conjunction displays. Although participants did not report experiencing a pop-out and their RTs were not as short as with pure single-feature displays, they showed some graded facilitation in RTs due to the very brief time period in which the target object was surrounded only by single-feature distractors. Olds et al. described this effect as a partial pop-out process assisting the difficult search process, and called it *search assistance*.

Thus, on the basis of these results of incremental display presentation, it appears that, rather than a pop-out or a no-pop-out signal's simply being output from a collection of individual feature maps to a more complex attentional search process, some form of graded or probabilistic information is continuously cascaded (see, e.g., McClelland, 1979) from the *parallel-like* feature maps to a *serial-like* search process. In one theoretical framework, it has been argued that information from the feature maps can "guide" the attention-based serial search. Rather than providing a single-stage account of visual search, this guided search model (see, e.g., Wolfe, 1994; Wolfe, Cave, & Franzel, 1989) challenges the one-way relationship between two independent processing stages, supporting the view that there is no sharp border between the parallel and the serial search strategies. Rather, the two types of processes may work concurrently to solve visual problems through parallel guidance of serially deployed attention.

On a much broader time scale than that used in Olds et al.'s (2000a, 2000b) experiments, Watson and Humphreys (1997, 2002) presented one set of colored distractors for a full second before adding to the display a set of different-colored distractors plus the target. They found that participants seemed to remember which distractors had already been discarded as nontargets during that initial second and thus were able to locate the target very efficiently. Watson and Humphreys suggested that this initial visual marking of distractors was a time-consuming, top-down inhibition process that allowed rejected distractors to not be rechecked. A different set of studies indicates that prioritization of new objects may depend on whether or not the new elements are presented with an abrupt onset upon their appearance (Donk & Theeuwes, 2001; Donk & Verburg, 2004). Moreover, there is evidence that visual search may not be guided entirely by memory-driven processes (Horowitz & Wolfe, 2003). Despite these objections, the visual marking phenomenon provides direct evidence that incremental display presentation increases search efficiency.

Overall, the reviewed studies point toward a continuum of search efficiency, suggesting that visual search may be a dynamic process of forming object representations out of feature information in real time. As such a representation is being formed, it communicates information to a search/decision process that waits for a sufficiently salient object to emerge. This process can be facilitated by at least two kinds of help—search assistance and visual marking—that seem to work by narrowing down the set of locations at which the target might be found. For example, in the experiments conducted by Olds et al. (2000a, 2000b) and Watson and Humphreys (1997, 2002), the visual display was incrementally delivered, providing information about locations that no longer needed to be considered and narrowing the search through a restriction in the number of relevant locations.

Another type of assistance to visual search comes from the identification of relevant versus irrelevant feature identities of the target item. Along this line, Spivey, Tyler, Eberhard, and Tanenhaus (2001) applied a manipulation involving incremental delivery of the target's identity (rather than incremental delivery of the search display). They showed that when linguistic information for target features is presented incrementally and concurrently with the visual display (auditory/visual [A/V]-concurrent condition), RTs were considerably less sensitive to the number of distractors. They argued that the notable improvement in search efficiency could be interpreted as a result of an

early and fluid interaction between linguistic processing and visual perception. That is, if a spoken phrase such as *the red vertical* is processed incrementally (see Altmann & Kamide, 1999; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995) and there is a rapid integration between partial linguistic information and visual representations, then the search process might be able to subtly enhance the salience of the subset of items exhibiting the target feature first mentioned (i.e., redness) in parallel, immediately after that adjective is heard, and then, upon hearing the second adjective (*vertical*), substantially enhance the salience of the single vertical object in that subset. Such a process would produce a graded improvement in search efficiency and, thus, quantitatively reduce the RT × set size slope.

It should be noted that the rapidity with which incremental linguistic input can modulate search efficiency is not unlimited. (Finding the limitations of this phenomenon is exactly what will allow the development of a more explicit mechanistic description of how concurrent linguistic input interfaces with the visual search process.) When Gibson, Eberhard, and Bryant (2005) replicated Spivey et al.'s (2001) experimental design and compared speaking rates of 3 and 4.8 syllables/sec, the improvement in search efficiency exerted by concurrent linguistic input disappeared with the faster speech.[1] In the faster speech condition, it appears as though the second adjective may be processed before the first adjective has had time to sufficiently enhance the salience of its subset of objects (and suppress the salience of the remaining objects). As a result, participants' performance is consistent with their conducting a standard inefficient conjunction search, just as in the control condition. Thus, Gibson et al.'s results are compatible with the claim that when speaking rate is somewhat slow, participants may be able to use each adjective the moment it comes in to constrain the visual search process. However, when the spoken query is delivered at a faster rate, the visual search process may be less capable of immediately utilizing the first adjective to enhance the salience of the subset of objects that share that first-mentioned feature.

Subset search strategies have been proposed to play a role in visual search under circumstances of explicitly guided search among a subset of the items shown (see, e.g., Egeth, Virzi, & Garbart, 1984; Kaptein, Theeuwes, & van der Heijden, 1995). For example, in Egeth et al., participants searched for a red "O" in a display of red "N"s and black "O"s. Varied ratios of red and black items were used in the display. If the number of black items was varied while the number of red items was held constant, search functions had flat slopes, suggesting that the participants were able to restrict their search to the subset of red items. However, in the same study, participants were explicitly instructed in advance to search through the subset of "O"s or through the subset of red items. Therefore, it is not clear whether or not the participants would have spontaneously adopted such a strategy of selecting the smaller subset of items sharing one of the target characteristics. One possibility is that concurrent integration of linguistic and visual information, as in Spivey et al. (2001), automatically triggers a similar kind of subset search strategy, explaining the shallow slopes in the A/V-concurrent condition.

The twofold purpose of the present work is to generalize Spivey et al.'s (2001) findings to a wider set of methodological circumstances that will test some of these strategy explanations, and to tease apart some potential mechanistic constraints for developing an explicit model of this interaction between vision and language. In Spivey et al.'s (2001) original study, the experiment involved separate blocks of an A/V-concurrent condition versus an auditory-then-visual control condition. Whenever a visual attention experiment has blocked trials, there is concern that practice may somehow be more effective in one condition than in the other or that participants may develop a deliberate strategy or mental set in one condition and not in the other (see, e.g., Müller, Heller, & Ziegler, 1995; Posner, Snyder, & Davidson, 1980; Wolfe, Butcher, Lee, & Hyle, 2003). In contrast, when trials from different conditions are mixed, it is less likely that independent strategies are used for different types of trials. In order to rule out some strategy explanations for Spivey et al.'s (2001) basic results, the present study explores the same basic paradigm that they used, but with different experimental designs. In our first experiment, we replicate the results when A/V-concurrent and control conditions are presented in a single block of randomly mixed trials, as opposed to the blocked condition design tested in the original study. In our second experiment, we extend this mixed-trials experimental design to conditions in which the order of adjectives (color and orientation) in the spoken query varies. In Experiment 3, we extend the design to a triple-conjunction task in which the target object is defined by three features (color, orientation, and size), to test whether each adjective can trigger its own nested parallel search.

Although the results of these first three experiments are consistent with an online influence of linguistic input on visual search, it could be argued that the process that causes the improved search efficiency in the A/V-concurrent condition is independent of the linguistic cues. This is because, in theory, participants can do this task without actually listening to the informative cue. For example, given a display with several red horizontal and green vertical items, distractors implicitly specify the target identity. If the target was a red vertical, distractors were invariably red horizontals and green verticals. Thus, after a number of trials a participant could realize that they are looking for a unique conjunction of color and orientation. Hypothetically, if the participant decides to attend to red, he could simply check if there is an "oddly" oriented item among the red items, and if the search among red items fails, the process would be repeated for the green items. It has been proposed that visual search can be efficiently guided on the basis of inferences of this type (Wolfe, 1992). Thus, this hypothetical odd-one-out search strategy could be producing shallow slopes in the A/V-

concurrent condition without the need for any interaction between language and vision. In order to test this possibility, we conducted a fourth experiment, in which a neutral, uninformative cue was delivered in the instruction, forcing participants to look for a unique conjunction without knowing its color or orientation. The results demonstrate that when the instruction contains an uninformative cue, the search is significantly less efficient than that in the A/V-concurrent condition with informative cues, suggesting that participants are indeed using the concurrently delivered spoken adjectives to search for the target.

Overall, this series of experiments excludes several alternative explanations of our observed linguistic assistance of visual search efficiency and takes important steps toward identifying the constraints and mechanism(s) whereby spoken adjectives, presented concurrently with the visual display, exert their influence on the search process. Experiment 1 rules out strategy effects that may have been induced by blocked trials in the previous experiments in the literature. Experiment 2 suggests that the order of adjectives may affect the degree of search efficiency elicited. Experiment 3 demonstrates that the improvement in search efficiency can also occur with a series of three adjectives. Experiment 4 excludes the possibility that participants ignore the adjectives in the spoken instruction and simply conduct an efficient odd-one-out search. Finally, simulations with a localist attractor network demonstrate how a parallel search process could allow the salience of the target to gradually emerge over time to produce linear RT functions. These simulations also show how the onset of the first adjective in the AV-concurrent condition could begin a gradual enhancement of the objects exhibiting that feature, and the second adjective could elicit a near-pop-out effect for the target object amidst the salience-enhanced objects that share the first-mentioned feature. Together, these results and simulations provide compelling support for the claim that visual search can be influenced in real time by continuous spoken language comprehension. They also contribute the initial steps toward developing and constraining a mechanistic account of how linguistic input exerts this influence.

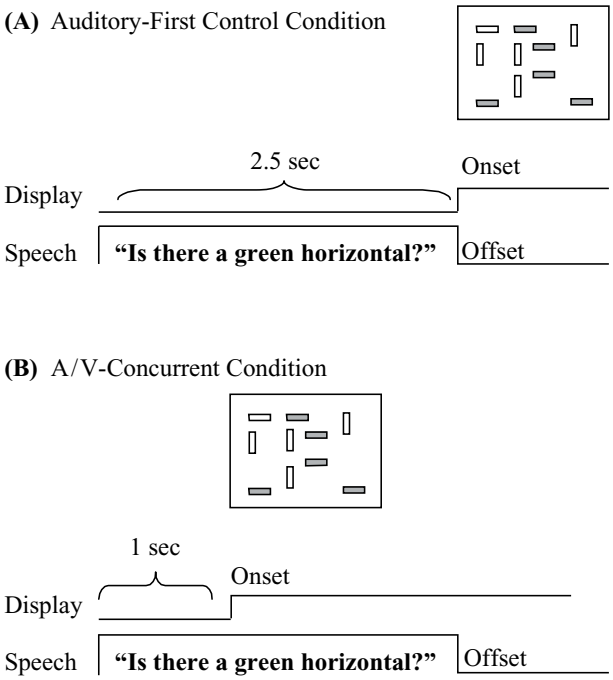

Figure 1. Outline of the auditory and visual stimuli. In the auditory-first control condition (A), the onset of the display coincided with the offset of the spoken target query. In the auditory/visual (A/V)-concurrent condition (B), the onset of the visual display coincided with the onset of the first target-feature word in the spoken query. The example displays show a target-present trial with a set size of 10. Dark bars represent red bars, and white bars represent green ones. In all conditions, reaction times were recorded from the onset of the visual display.

EXPERIMENT 1

In this experiment, we replicated the design of Spivey et al.'s (2001) Experiment 1, except that our design was mixed instead of blocked, and the A/V-concurrent and control trials were randomly interspersed.

Method

Participants. Eighteen Cornell undergraduate students participated in this experiment, receiving extra credit in psychology courses. All the participants had normal or corrected-to-normal vision and normal color perception.

Stimuli and Procedure. The experiment was composed of two types of trials—A/V-concurrent trials and auditory-first control trials—presented in a random order within one block of 192 trials. In the auditory-first control condition, the participant received the complete auditory target query (e.g., "Is there a green horizontal?") before the visual presentation of the target display. In the A/V-concurrent condition, the same spoken query was presented in such a way that the onset of the first adjective coincided with the onset of the visual display (see Figure 1). Each condition contained 96 trials within the mixed block. The same 96 visual displays and four prerecorded target queries were used in both conditions.

The participants were allowed to begin each trial when they were ready. They were instructed to respond as quickly and accurately as possible to the questions, delivered in the form of digitized speech file, by pressing the "yes" and "no" buttons for target present and target absent, respectively. Throughout all our conditions and all our experiments, the RTs for this response were measured from the onset of the visual display to the moment the response button was pressed. An initial fixation cross preceded the onset of the visual display in order to direct the participant's gaze to the central region of the display. Each stimulus bar subtended 2.8º × 0.4º of visual angle, and neighboring bars were separated from one another by an average of 2º of visual angle. The white background had a luminance of 68.4 cd/m^2. The red bars had a luminance of 25.3 cd/m^2 and CMYK color parameters of {0,80,100,0}. The green bars had a luminance of 31.6 cd/m^2 and CMYK color parameters of {100,40,80,40}. The bars appeared in a random arrangement, constrained by an invisible grid positioned centrally on the screen. Set sizes for the visual displays were 5, 10, 15, and 20. Equal numbers of trials with each type of target were randomly distributed across the session. On any given trial, roughly half of the distractors exhibited one of the target features and the other half exhibited the other target feature. The recorded voice was that of the same female speaker on all trials, and all the speech files contained the 1-sec preamble recording, "Is there a . . ." spliced onto the beginning of each of the four target queries ("red vertical?," "red horizontal?," "green vertical?," and

"green horizontal?"). Throughout all the experiments of the present study, the adjectives themselves were spoken at an average rate of 3 syllables/sec.

Results and Discussion

Mean accuracy was 95% and did not differ significantly across conditions. Figure 2 shows the RT × set size functions for target-present and target-absent trials in the A/V-concurrent condition and the auditory-first control condition. The best-fit linear equations and the corresponding r^2 values indicate the proportion of variance accounted for by the linear regression. Note that when the RT × set size slopes are quite shallow (as in the A/V-concurrent condition), it is common for their linear regression fits to be less than robust (see, e.g., Treisman & Gelade, 1980). In the target-absent trials, both the auditory-first and A/V-concurrent conditions presented highly linear functions of RT × set size, as is indicated by the r^2 values. However, in the target-present trials, the RT × set size function for the A/V-concurrent condition was notably less linear than the function for the control condition. Since the complete auditory notification of the target's identity was delayed by approximately 1.5 sec in the A/V-concurrent condition in comparison with the auditory-first condition, a longer mean RT is observed in the former condition. However, because spoken word recognition is incremental, the participants were able to process the information concurrently with the target feature notification, resulting in a delay in overall RT of only about 600 msec rather than 1.5 sec.

The most important finding for the purpose of our inquiry was the significant interaction between condition and set size in the target-present trials. A repeated measures ANOVA revealed that the effect of set size was more pronounced (i.e., had a steeper slope) in the auditory-first control condition than in the A/V-concurrent condition [$F(3,48) = 4.36$, $p < .01$]. Even though the visual displays were identical in both conditions and the same speech files were merely shifted in time by about 1.5 sec, the A/V-concurrent condition produced shallower slopes (reflecting higher visual search efficiency) than did the auditory-first control condition (see Figure 2). Interestingly, during debriefing several participants spontaneously reported being unaware that half of the trials had one kind of timing for their spoken target query deliveries and the other half had another kind of timing. They claimed that they had not noticed the difference between the A/V-concurrent trials and the auditory-first trials.

Although the regression fit for the A/V-concurrent target-present trials was rather weak due to the function's becoming slightly nonlinear as it approaches flatness, it is worthwhile to conduct comparisons of RT × set size slopes across conditions (calculated within subjects and averaged across them) just to provide an additional test of whether or not search efficiency is indeed improved in the A/V-concurrent condition. To specifically test whether the participants' RT × set size slopes were significantly shallower in the A/V-concurrent condition, the participants' individual set size slopes from the two conditions were

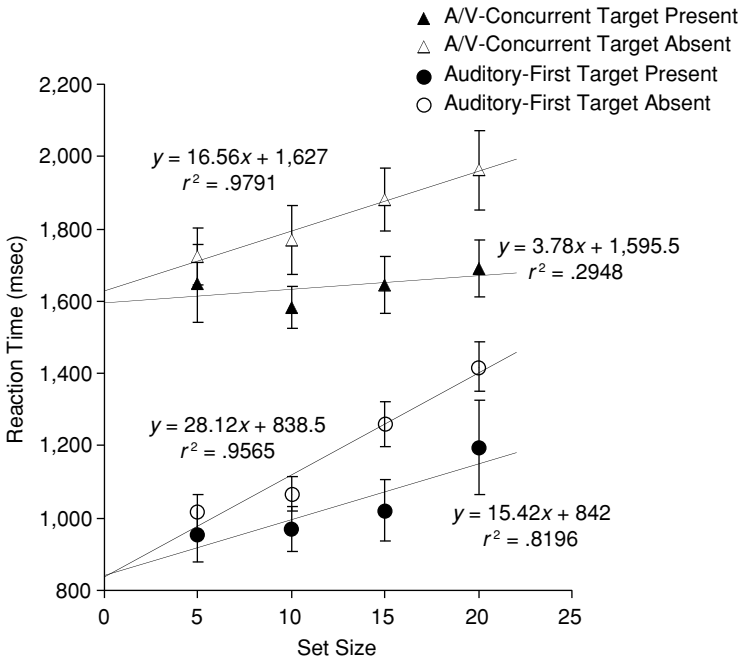

Figure 2. Results from Experiment 1, shown separately for target-present (filled symbols) and target-absent (open symbols) trials for both the auditory-first control condition (circles) and the A/V-concurrent condition (triangles). Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

compared via paired *t* tests. The results indicated shallower slopes for the A/V-concurrent condition than for the auditory-first condition in target-present trials [3.78 vs. 15.42 msec/item; $t(16) = 2.09, p < .05$] and in the target-absent trials [16.56 vs. 28.12 msec/item; $t(16) = 3.39, p < .05$]. The nearly 2:1 ratio between target-absent and target-present trials in the auditory-first condition (28.12 vs. 16.56 msec/item) is consistent with a standard serial search account. Curiously, this ratio is substantially higher in the A/V-concurrent condition than in the auditory-first condition (4:1, with 16.56 vs. 3.78 msec/item), a result that is not expected under the hypothesis that the participants were engaged in a serial search among a subset of the display items. A closer look at the slope values reveals that the approximately 4:1 slope ratio found in the A/V-concurrent condition did not result from an inefficient search in the target-absent condition but, rather, from an unusually efficient search in the target-present condition. In the present experiment, the slope for the target-present condition is several milliseconds per item shallower than in the blocked-trials version of the same experiment (Spivey et al., 2001, Experiment 1), and the slope for the target-absent condition is also shallower by a few milliseconds per item. However, this turns the previous slope ratio from about 3:1 to 4:1. Thus, it may be the case that, no matter how efficient the A/V-concurrent target-present trials become, the target-absent trials are simply unable to exhibit slopes of less than about 15 msec/item. This may suggest that the cognitive process that determines a target-absent response may be something substantially different than the process that determines a target-present response (cf. Chun & Wolfe, 1996), rather than both being the same (serial search) process that simply takes twice as long in the target-absent condition.

Consistent with previous results reported in Spivey et al. (2001), we found that the search process showed dramatically improved search efficiency in the A/V-concurrent condition. The mean slope for A/V-concurrent target-present trials is in the range of standard single-feature pop-out conditions (see, e.g., Treisman, 1988; Treisman & Gelade, 1980). In the search process in the auditory-first condition, both features influence search simultaneously, thus causing each distractor to exhibit an equal partial match (.5 similarity) with the target parameters and thereby reducing the efficiency with which the target object's salience can come to the foreground. However, in the A/V-concurrent condition it appears that the incremental nature of the speech input allows the search process to begin when only a single feature of the target identity has been heard. Thus, the salience of objects exhibiting the first-mentioned feature can be enhanced whereas that of the other objects is suppressed. Then, when the second adjective is heard, it can significantly enhance only the salience of the target object because the other objects it would have enhanced have been suppressed. The fact that we replicated Spivey et al.'s (2001) findings using a mixed-block design indicates that the effect is not the result of differential strategies in the A/V-concurrent and control conditions.

EXPERIMENT 2

In this experiment, we explored whether or not this real-time interaction between incremental linguistic processing and visual search is still present when one increases the complexity of the linguistic component of the task by making the target queries more varied. So far, in each experiment that has demonstrated this improved search efficiency, only four types of queries have been used. Therefore, it is possible that participants in the previous experiments may have been employing listening strategies (e.g., just discriminating the word *red* from the word *green*) that are not representative of everyday language comprehension.

This expanded version of the previous experiments had eight different queries and more trials (but fewer trials per condition). Half of the spoken target queries delivered the color adjective before the orientation adjective, whereas the other half of the queries delivered the orientation adjective before the color adjective. In fact, in Experiment 2 of Spivey et al. (2001), in which the orientation adjective preceded the color adjective, the A/V-concurrent condition reduced the RT × set size slope by a numerically smaller degree than in their Experiment 1 (in which the color adjective always preceded the orientation adjective). Moreover, the difference in slopes was only marginally significant for the target-absent trials of Spivey et al.'s (2001) Experiment 2. Therefore, it is possible that the fluidity with which concurrent spoken linguistic input influences visual search is affected by the adjective order being used.

Method

Participants. Fifty-six undergraduate students participated in the experiment, receiving extra credit in psychology courses. (More cells in the factorial design and fewer trials per cell led us to run more participants in this experiment than in the others.) All the participants had normal or corrected-to-normal vision and normal color perception.

Stimuli and Procedure. In this experiment, adjective order varied in the spoken instruction. Color-first and orientation-first query trials (e.g., "Is there a green vertical?" and "Is there a vertical green?," respectively) were randomly interspersed across conditions. The experiment included four types of trials (A/V concurrent with color-first query, A/V concurrent with orientation-first query, auditory first with color-first query, and auditory first with orientation-first query) presented in a random mixed order within one block of 256 trials. We used an expanded set of visual displays based on those from Experiment 1, and stimuli had the same timing (as presented in Figure 1). Prerecorded spoken queries (color first or orientation first for all four targets) were identical in both the auditory-first and A/V-concurrent conditions. Set sizes of objects comprising the visual displays were 5, 10, 15, and 20.

Results and Discussion

Mean accuracy was 96.5% and did not differ significantly across conditions. RT × set size functions for target-present and target-absent trials in the A/V-concurrent condition and the auditory-first control condition are shown in Figures 3 (color-first trials) and 4 (orientation-first trials). As was expected in the A/V-concurrent condition, be-

cause the participants had to wait until target features were spoken, they produced mean RTs approximately 800 msec longer for color-first trials and about 1,100 msec longer for orientation-first trials, in comparison with the corresponding auditory-first control conditions. The overall difference between adjective order trials resulted from the fact that the color adjective is a one-syllable word that is shorter in duration than the orientation adjective. As is indicated by the r^2 values in Figures 3 and 4, both the auditory-first and A/V-concurrent conditions obtained highly linear RT × set size functions.

A repeated measures ANOVA on the target-present trials, collapsed across adjective order, revealed a significant interaction between condition (A/V concurrent vs. auditory first) and set size. That is, the effect of set size was more pronounced in the auditory-first condition than in the A/V-concurrent condition for target-present trials [$F(3,165) = 2.76$, $p < .05$]. To specifically test whether or not the mean slope was significantly shallower in the A/V-concurrent condition, the participants' individual set size slopes from the two conditions were compared via paired *t* tests. First, we compared mean slopes for conditions collapsed across adjective order and found shallower slopes for the A/V-concurrent condition in both the target-present [$t(55) = 2.23$, $p < .05$] and target-absent [$t(55) = 3.9$, $p < .001$] trials. When considering only color-first trials (Figure 3), we found shallower slopes for the A/V-concurrent condition than for the auditory-first condition in the target-present [2.61 vs. 9.19 msec/item; $t(55) = 2.69$, $p < .001$] and target-absent [15.14 vs. 23.84 msec/item; $t(55) = 3.38$, $p < .001$] trials. The present:absent slope ratio in this A/V-concurrent condition is similar to that found in Experiment 1, and the auditory-first condition again produced a nearly 2:1 ratio, consistent with a standard serial search. The unusually shallow slope found for target-present trials in this A/V-concurrent condition suggests that the participants may have followed a search strategy that is actually more efficient than a serial search among half of the items in the display. Paired *t* test analyses in orientation-first trials (Figure 4) showed significantly shallower slopes for the A/V-concurrent than for the auditory-first condition in target-absent trials [17.59 vs. 23.07 msec/item; $t(55) = 2.07$, $p < .05$] but not in the target-present trials. Although the mean slope was numerically shallower in the A/V-concurrent than in the auditory-first condition for orientation-first queries in the target-present trials (9.10 vs. 13.00 msec/item), the difference was not significant, suggesting that the order of adjectives may be an important factor in eliciting the observed interaction between language processing and visual search.

The color-then-orientation target query introduced a more robust improvement of search efficiency than the orientation-then-color target query when delivered con-

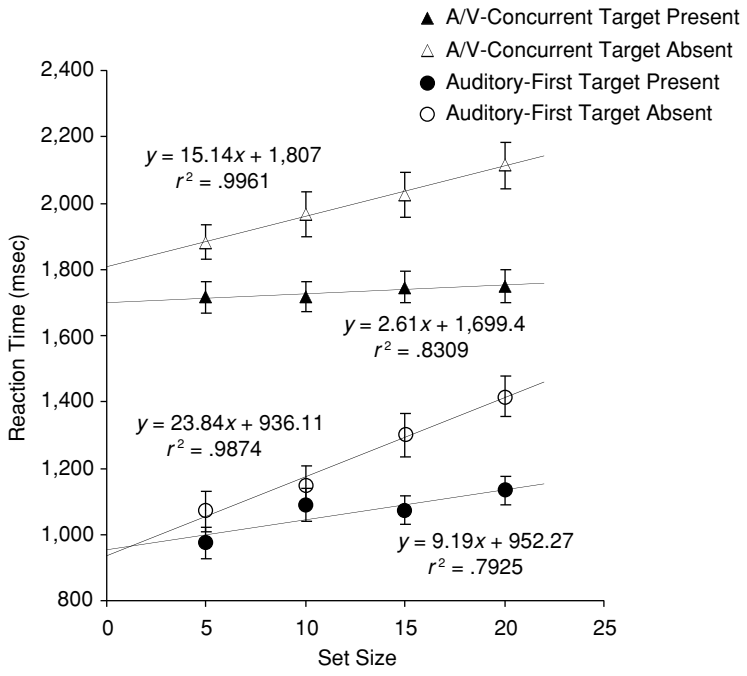

Figure 3. Results from Experiment 2, corresponding to the color-first trials. The spoken target queries in this experiment were in the form, "Is there a [color] [orientation]?" The results are shown separately for target-present (filled symbols) and target-absent (open symbols) trials for both the auditory-first control condition (circles) and the A/V-concurrent condition (triangles). Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

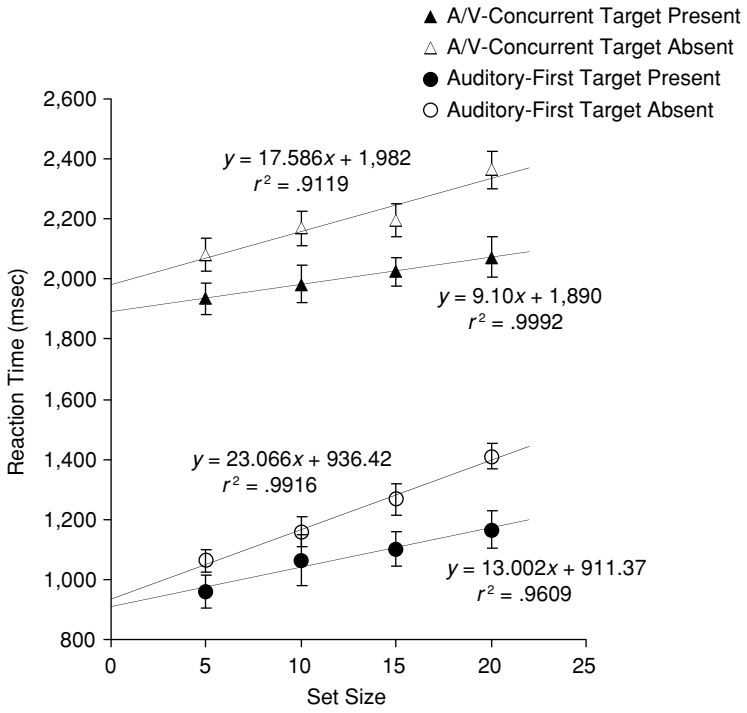

Figure 4. Results from Experiment 2, corresponding to the orientation-first trials. The spoken target queries in this experiment were in the form, "Is there a [orientation] [color]?" The results are shown separately for target-present (filled symbols) and target-absent (open symbols) trials for both the auditory-first control condition (circles) and the A/V-concurrent condition (triangles). Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

currently with the display. This asymmetry is consistent with the preferred order of feature type delivery observed in other studies. For example, Olds and Fockler (2004) conducted purely visual experiments with no linguistic input in which they found greater search assistance from single-feature previews that primed color first and then orientation than from those that primed orientation first and then color, suggesting that the cause of the asymmetry may reside in visual discrimination or localization of color versus visual discrimination or localization of orientation. Moreover, Boucart and Humphreys (1997) reported that in a matching task attention could be efficiently applied to color without influence from semantic information regarding the objects, but semantic information did interfere with attention to orientation. Thus, our asymmetry in results with the different adjective orders may be due to differences in the overall visual salience of color versus orientation.

EXPERIMENT 3

In Experiment 3, we explored whether or not the increased efficiency obtained in the A/V-concurrent condition is also present when the target is defined by the conjunction of three features: color, orientation, and size (cf. Quinlan & Humphreys, 1987). If search efficiency is best described as a continuum of efficiency rather than as a dichotomy between parallel and serial mechanisms, then the improvement that we have been observing in the A/V-concurrent condition may be characterized as a successive graded enhancement of feature salience resulting from the sequential delivery of each target adjective. The comprehension of the adjective spoken first (during viewing of the target display) may allow the search to begin in a fashion that gradually highlights the subset of objects with the corresponding feature in such a way that, when the second adjective arrives, the target can be quickly identified from amidst that subset. If there is merit in this description of how concurrent linguistic delivery of target identity improves search efficiency, then we would expect an A/V-concurrent triple-conjunction search to produce an improvement in search efficiency similar to that of the auditory-first control condition. Essentially, one could expect the first adjective to begin the gradual enhancement of the many objects exhibiting its feature, the second adjective to begin the gradual enhancement of the several objects within that subset that exhibit its feature, and the third adjective to quickly single out the one object in that sub-subset that exhibits its feature.

Triple-conjunction search displays in which each distractor shares only one feature with the target tend to elicit shallower RT × set size functions than standard conjunction search displays (Wolfe et al., 1989). However, triple-

conjunction search displays in which each distractor shares two features with the target generally elicit *steeper* RT × set size functions than standard conjunction search displays. Since steeper slopes for the control condition allow more room for observing an improvement in search efficiency (and shallower slopes risk a floor effect that may prevent improved efficiency from being statistically discernable), we chose to test the effect of concurrent incremental linguistic target delivery on triple-conjunction search displays that had distractors sharing two features with the target.

Method

Participants. Sixteen undergraduate students participated in the experiment, receiving extra credit in psychology courses. All the participants had normal or corrected-to-normal vision and normal color perception.

Stimuli and Procedure. In this experiment, the adjective order in the spoken query was constant across trials: color first, then orientation, and finally size (i.e., "Is there a [red/green] [vertical/horizontal] [large/small]?"). In the series of trials, each of the eight types of targets and each of the eight target queries was used 32 times. A/V-concurrent and auditory-first conditions were randomly mixed across one block of 256 trials. The recorded voice was that of the same female speaker as in Experiments 1 and 2. Each speech file contained an identical 1-sec preamble ("Is there a . . ." spliced onto the beginning of each of the eight target queries ("red vertical large?," "green horizontal small?," etc.). Timing of stimulus presentation was the same as in Experiments 1 and 2. Objects comprising the visual display appeared in a grid-like arrangement positioned centrally on the screen, as in Experiments 1 and 2. So that each target-present display could have an equal number of the three types of distractors (each sharing two of the target's features), we used set sizes of 7, 13, 19, and 25.

Results and Discussion

Mean accuracy was 92% and did not differ across conditions. Figure 5 shows the RT × set size functions for target-present and target-absent trials in the A/V-concurrent condition and the auditory-first control condition. For reasons similar to those in Experiments 1 and 2, the participants exhibited a mean RT approximately 1,250 msec longer in the A/V-concurrent condition than in the auditory-first control condition.

As in Experiments 1 and 2, a repeated measures ANOVA on the target-present trials revealed a significant interaction between A/V-concurrent versus auditory-first conditions and set size [$F(3,45) = 13.48$, $p < .001$]. A paired t test analysis of the participants' individual set size slopes indicated significantly shallower slopes for the A/V-concurrent than for the auditory-first condition in both target-present [10.93 vs. 23.35 msec/item; $t(15) = 3.51$, $p < .01$] and target-absent [40.1 vs. 65.28 msec; $t(15) = 5.06$, $p < .001$] trials. As before, these results show absent:present slope ratios that are substantially greater than the 2:1 ratio predicted by a serial object-by-object search, further suggesting that target-present responses may involve something other than a serial self-terminating

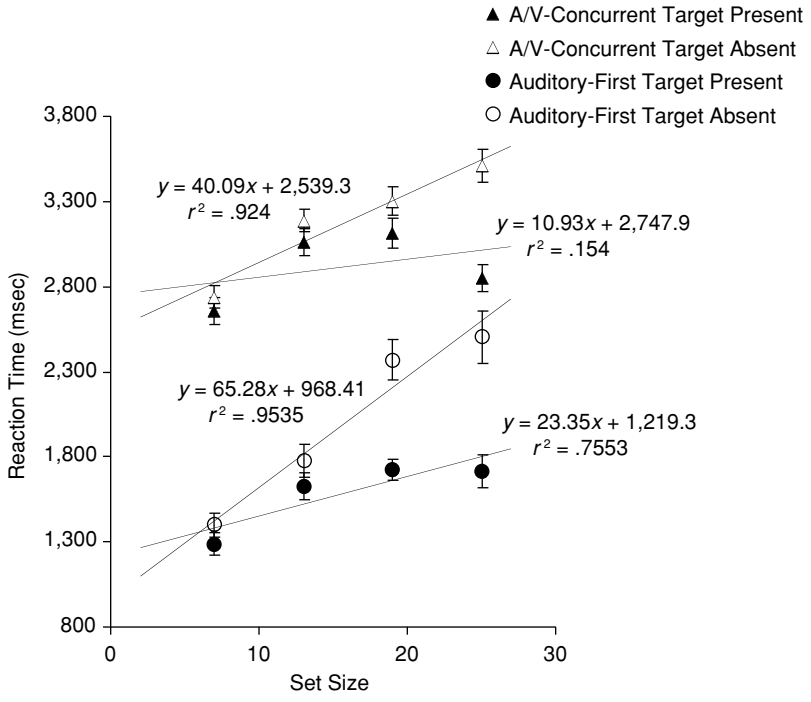

Figure 5. Results from Experiment 3. The spoken target queries in this experiment were in the form, "Is there a [color] [orientation] [size]?" The results are shown separately for target-present (filled symbols) and target-absent (open symbols) trials for both the auditory-first control condition (circles) and the A/V-concurrent condition (triangles). Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

search of the display, whereas target-absent responses may involve something other than a serial exhaustive search of the display.

The fact that the slopes are shallower for the A/V-concurrent condition than for the auditory-first control condition indicates that incremental linguistic processes do improve the efficiency of visual search even when it is an unusually difficult triple-conjunction search. In fact, the amount of reduction in slope induced by playing the adjectives *after* the display onset instead of *before* it is about 12 msec/item for target-present trials, which is quite close to the amount of reduction seen in Experiment 1 and in Spivey et al. (2001). The presence of this quantitative improvement in search efficiency (instead of a dichotomous shifting between steep and flat functions) is consistent with the idea that incremental delivery of the target identity via linguistic input that is concurrent with the display improves search efficiency in a graded fashion. In terms of a biased competition account (Desimone & Duncan, 1995), what may be happening is that the first adjective increases the salience of two thirds of the objects, but since they compete against one another they cannot become especially active (see Figure 6). Meanwhile, the objects not exhibiting the feature corresponding to the first adjective are gradually suppressed in activation. Then, the second adjective increases the salience of half of the objects in the enhanced subset, but those objects compete against one another as well. Finally, the third adjective increases the salience of only one object amidst that sub-subset. Since it has no significant competitors, it wins quickly and efficiently.

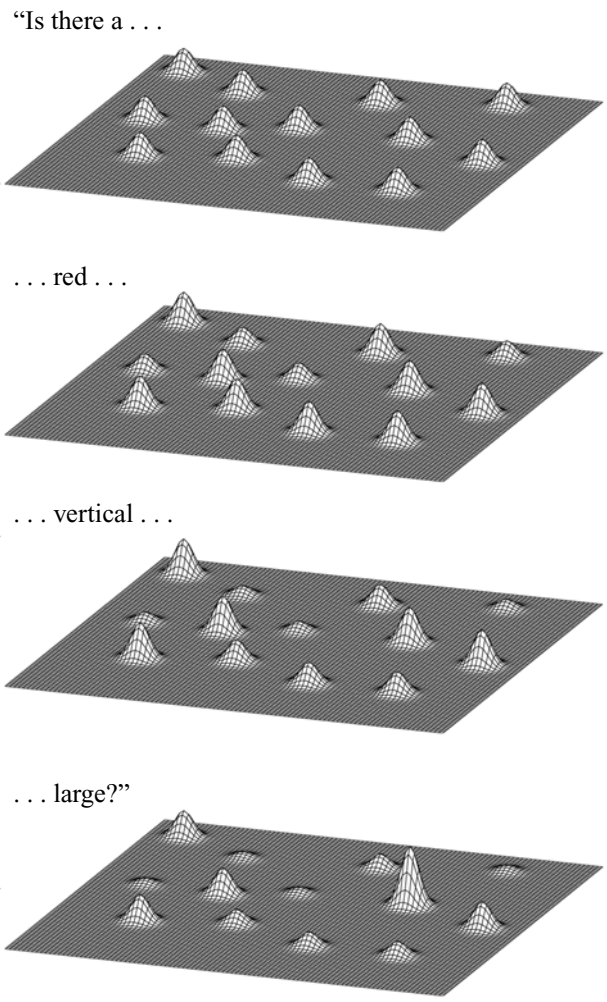

Figure 6. A schematic depiction of a salience map with activations for objects changing in parallel as each adjective is processed in an A/V-concurrent triple-conjunction search, with a set size of 13. When "red" is heard, the 9 red bars are enhanced and the 4 green vertical large bars are suppressed. When "vertical" is heard, the 5 red vertical bars continue to increase in salience, whereas all others decrease. Finally, when "large" is heard, only the target object continues to rise dramatically, whereas all others drop.

EXPERIMENT 4

Experiment 4 is a control experiment designed to test whether participants in the A/V-concurrent condition are actually being guided by the linguistic cues or are instead following a subset search strategy guided by inferences about the target's uniqueness (see, e.g., Treisman & Sato, 1990; Wolfe, 1992). This experiment involved separate blocks for the A/V-concurrent condition, in which linguistic cues were used, and for a control A/V-concurrent condition, in which a neutral uninformative query ("Is there an odd one out?") was used. In the neutral control condition, observers are left to determine the identity of the target on their own. This kind of inference can work because the distractors implicitly specify the target's identity. For example, if the distractors are red horizontal and green vertical bars, then the target is either the sole vertically oriented bar among the red ones or the sole horizontally oriented bar among the green ones. A participant can, in principle, simply check whether or not the red objects have a uniquely oriented rectangle among them and, barring that, simply repeat the check for the green objects. Crucially, in both conditions in this experiment the verbal instruction was delivered concurrently with the visual display and with the same presentation timing. If visual search in the standard A/V-concurrent condition is actually guided merely by odd-one-out inferences regarding the target and its distractors (and if participants are essentially ignoring the spoken instruction), then we should expect comparably shallow slopes to be found in both conditions of this experiment for both target-present and target-absent cases. If, alternatively, linguistic interpretation of the spoken adjectives is instrumental in directing visual search, then we should find shallower slopes in the standard A/V-concurrent condition than in the uninformative neutral cue condition.

Method

Participants. Thirty Cornell undergraduate students participated in this experiment, receiving extra credit in psychology courses. All

the participants had normal or corrected-to-normal vision and normal color perception.

Stimuli and Procedure. In this experiment, there were two blocks of two trial types each. One of the blocks contained the A/V-concurrent informative cue (A/V-CIC) trials, and the other block contained the A/V-concurrent uninformative cue (A/V-CUC) control trials. Each block contained 96 randomly ordered trials. The two blocks were conducted in one order for half of the participants and in the alternate order for the other half. In the A/V-CIC condition, both the spoken instructions and the timing of stimulus presentation were identical to those in the A/V-concurrent condition in Experiment 1. In the A/V-CUC control condition, the participants received the following auditory target query containing the uninformative verbal cue, "Is there an odd one out?" The timing of the instruction and stimulus presentation was such that the onset of the word *odd* coincided with the onset of the visual display, as in the A/V-CIC condition. The recorded voice was that of the same female speaker as in the A/V-CIC condition and in Experiments 1, 2, and 3. The 96 visual displays used in Experiment 1 were used in both conditions of the present experiment.

At the beginning of the experiment, both conditions were explained to the participants, who were exposed to four practice trials containing two examples from each condition. Before being exposed to the practice trials, the participants were explicitly told that an item described as an "odd one out" would be an object that looks like no other object in the display. The participants were instructed that if an "odd one out" was present in the display, it would have a unique conjunction of two features, such as being the only red bar that is vertical or the only green bar that is horizontal. They were instructed to respond as quickly and accurately as possible to the question, delivered via a digitized speech file, by pressing the "yes" and "no" buttons for target present and target absent, respectively. As in Experiments 1 and 2, the set sizes for the visual displays were 5, 10, 15, and 20.

Results and Discussion

Mean accuracy was 97% and did not differ significantly across conditions. Figure 7 shows the RT × set size functions for target-present and target-absent trials in the A/V-CIC condition and the A/V-CUC control condition. The best-fit linear equations and the corresponding r^2 values indicate the proportion of variance accounted for by the linear regression.

When the spoken query was the uninformative cue (i.e., "Is there an odd one out?"), the participants could begin their odd-one-out search at the moment of the onset of the word *odd*. In contrast, when the spoken query was an informative cue (e.g., "Is there a green vertical?"), it appears that the participants were indeed paying attention to the sequentially delivered adjectives and listening at least until they heard the onset of the second adjective. As a result, the *y* intercept of the RT × set size function is slightly higher in the informative than in the uninformative cue condition.

The most important observation in this control experiment was the significant interaction between condition and set size in the target-present trials. A repeated measures ANOVA revealed that the effect of set size was more pronounced (i.e., it had a steeper slope) in the A/V-CUC con-

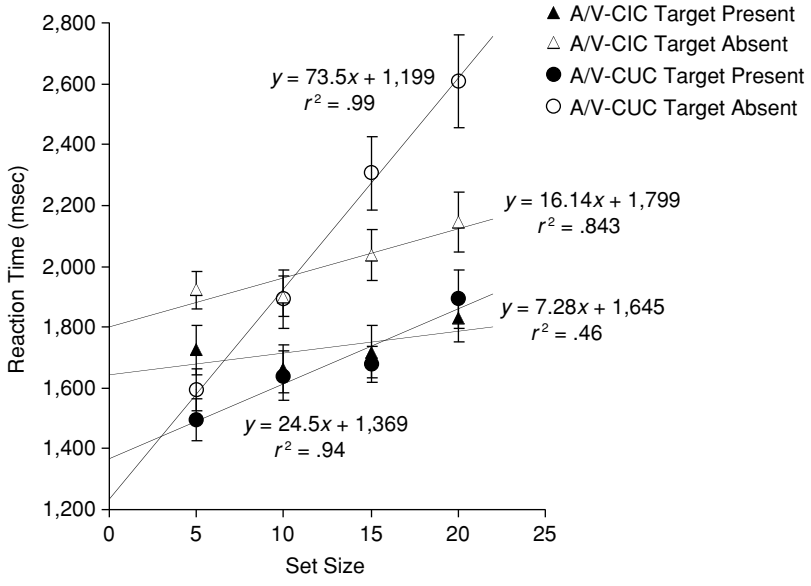

Figure 7. Results from Experiment 4. The spoken target queries in this experiment (all delivered concurrently with the visual display) were in the form of "Is there a [color] [orientation]?" in the concurrent informative cue (CIC) condition and in the form of "Is there an odd one out?" in the concurrent uninformative cue (CUC) control condition. The results are shown separately for target-present (filled symbols) and target-absent (open symbols) trials for both the control condition (circles) and the informative condition (triangles). Each line is accompanied by the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean.

trol condition than in the A/V-CIC condition [$F(3,87) = 5.85$, $p < .001$]. To specifically test whether or not the participants' RT × set size slopes were significantly shallower in the A/V-CIC condition than in the A/V-CUC control condition, the participants' individual set size slopes from the two conditions were compared via paired *t* tests. The results indicated shallower slopes for the A/V-CIC than for the A/V-CUC condition in target-present trials [7.3 vs. 24.8 msec/item; $t(29) = 3.19$, $p < .004$] and in the target-absent trials [16.1 vs. 73.5 msec/item; $t(29) = 7.95$, $p < .001$].

Curiously, the absent:present slope ratio for the A/V-CIC condition here is once again near the 2:1 range, as was obtained in Spivey et al.'s (2001) Experiment 2. Overall, the A/V-concurrent condition across five different standard-conjunction experiments has produced target-absent slopes that consistently range from 15 to 23 msec/item, and their corresponding target-present slopes have ranged from 3 to 9 msec/item. The substantial difference in absent:present slope ratios at the low and high ends of these ranges suggests that target-absent responses may not be simply the result of the search process's sequentially exhausting the entire set of possible targets. Rather, quantitative simulations (Chun & Wolfe, 1996) demonstrate that a visual search process may be terminated before all objects have been checked for their match to a target template, or that participants may even develop expectations of how long it *should* take to find the target and then terminate the search when that duration has elapsed. Thus, there is likely to be a number of circumstances under which the standard 2:1 absent:present slope ratio predicted by the original feature integration theory (Treisman & Gelade, 1980) will not obtain.

The steep slope values found in the A/V-CUC control condition were consistent with an inefficient conjunction search. These results clearly indicate that the spoken adjectives in the A/V-CIC condition provide an advantage over uninformative cues. Thus, the reduced slopes found across several experiments are better interpreted as the result of incremental linguistic input's assisting the search process by enhancing the salience of relevant objects as each adjective is heard, rather than as the result of a strategy of looking for any unique conjunction in the display.

GENERAL DISCUSSION

In a variety of different experimental designs (Experiments 1–3), the present results indicate that there is a significant improvement in visual search efficiency when the identity of the conjunction target is delivered incrementally via a spoken target query *while the stimulus display is visible*, rather than prior to stimulus onset. The underlying explanation could be that, in a dual conjunction task, as the linguistic information is processed continuously over time, the search process can enhance the salience of the subset of items sharing the feature mentioned first and suppress the salience of the other objects. Then, when the feature mentioned second is processed, its resulting enhancement of salience is effective only on the target object (since the other objects exhibiting that feature have been suppressed). In order to do that, the brain must be able to seamlessly cross-index partial linguistic representations (e.g., the first adjective of a phrase) with partial visual representations (of a cluttered visual display). In Experiment 4, we explored whether or not such a subset search strategy could result from a spontaneous search for the "odd one out" among a subset of the items in the A/V-concurrent condition. Such a strategy could be triggered by the initial delay between the visual display onset and the complete delivery of the information that identifies the target. However, the reliable difference between slopes across conditions in Experiment 4 indicates that participants follow different search strategies when they hear informative versus uninformative cues in the A/V-concurrent condition. Thus, the results suggest that when cues are informative, linguistic signals interact with the visual search process in a way that produces a quantitative improvement in search efficiency.

These results are in line with a rapidly growing body of work providing relevant behavioral experimental evidence of functional interaction between visual processing and linguistic processing. The vast majority of these studies demonstrate the influence of visual processes in real-time language comprehension. An example is the fact that visual and affordance-based contexts have an immediate influence on the online resolution of temporary ambiguities in spoken word recognition (Allopenna, Magnuson, & Tanenhaus, 1998; Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Spivey-Knowlton, Tanenhaus, Eberhard, & Sedivy, 1998) and in syntactic parsing (Chambers, Magnuson, & Tanenhaus, 2004; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Just as that literature demonstrates that visual input can influence real-time spoken language comprehension processes, the present results support the contention that linguistic input can influence real-time visual search processes.

A potential mechanism by which this incremental linguistic input could be influencing the search process is easily implemented in a simple localist attractor network simulation of RTs during visual search (Spivey & Dale, 2004). Whereas previous implementations of Desimone and Duncan's (1995) biased competition framework have focused at the level of firing rates of individual neurons (Reynolds & Desimone, 2001; Spratling & Johnson, 2004), Spivey and Dale's simulation of visual search RTs abstracted this framework to the level of functionally unitized population codes representing objects that compete against one another. In the present implementation of this model, one feature vector of 100 nodes registers "targethood" on the basis of redness (positive activation) and nonredness (zero activation) for up to 100 objects. Another 100-node feature vector registers "targethood" on the basis of verticalness and nonverticalness. Finally, a 100-node integration vector registers each object's likelihood of being the target. At the beginning of the simulation, the initial activation of each node in a feature vector is $1/N$, where N is the number of nodes in the vector. Input

to these feature vectors, resulting from hearing "red" and "vertical," involves multiplying each node by 1 if the object exhibits the relevant feature and by 0 if the object does not exhibit the relevant feature. Feature nodes that do not correspond to any object, as when the set size is less than 100, also receive 0 for their multiplicative input. When the attractor network begins its settling process, each time step begins with each feature vector being normalized to a sum total of 1, causing them to look a bit like probability distributions. (See Equations 1A and 1B, where $\mathbf{R}_{n,t}$ is the redness vector at time t and $\mathbf{V}_{n,t}$ is the verticalness vector at time t, respectively.) These vectors are then noncumulatively averaged at an integration layer ($\mathbf{I}_{n,t}$; see Equation 2). In this normalized recurrence competition algorithm (Spivey & Dale, 2004), the integration layer then sends point-wise multiplicative cumulative feedback to each of the feature vectors, biasing them in a fashion that exerts a little cross talk from the other feature vector (Equations 3A and 3B).

$$\mathbf{R}_{n,t} = \mathbf{R}_{n,t} \,/\, \sum_{n}\left(\mathbf{R}_{n,t}\right) \tag{1A}$$

$$\mathbf{V}_{n,t} = \mathbf{V}_{n,t} \,/\, \sum_{n}\left(\mathbf{V}_{n,t}\right) \tag{1B}$$

$$\mathbf{I}_{n,t} = .5 * \mathbf{R}_{n,t} + .5 * \mathbf{V}_{n,t} \tag{2}$$

$$\mathbf{R}_{n,t+1} = \mathbf{R}_{n,t} + \mathbf{I}_{n,t} * \mathbf{R}_{n,t} \tag{3A}$$

$$\mathbf{V}_{n,t+1} = \mathbf{V}_{n,t} + \mathbf{I}_{n,t} * \mathbf{V}_{n,t} \tag{3B}$$

For each time step (treated as 20 msec), this normalization–integration–feedback cycle repeats until a node in the integration layer exceeds some criterion activation (.95 in this case), at which point the target has been found and a settling time (i.e., RT) is recorded. A constant of 800 msec is then added to this RT for perceptual registration and motor execution. (Target-absent trials are not simulated by this competition algorithm, since their termination is not likely the result of a representation's winning a competition process; see Chun & Wolfe, 1996.) Importantly, this model allows the targethood of each object (in the integration layer) to be updated and evaluated in parallel at each time step, rather than imposing a serial search of one object at a time. Nonetheless, it produces a strikingly linear increase in settling time as set size increases. When the redness and verticalness vectors received input at the same time, simulating the auditory-first control condition, the result was an RT × set size slope of 17.2 msec/item (see Figure 8).

In simulating the A/V-concurrent condition, the redness feature vector received its input first, and the network was allowed to pursue its settling process for 35 time steps before the verticalness feature vector received its input (the equivalent of 700 msec from the point at which the first adjective is recognizable to the point at which the second adjective is recognizable). Under these conditions, the RT × set size slope was reduced to 6.8 msec/item. Basically, as the uneven activation pattern in the redness (first-mentioned) vector gradually biases the verticalness vector toward the red objects, the nonred objects have less and less probabilistic activation available to them. Then, when the external input to the verticalness vector is activated (because the adjective *vertical* is now being heard), it is multiplied by several nonred vertical nodes whose activations are already so close to zero (reduced by two thirds, from .01 to .0031) that they are unable to compete effectively. In this way, the model essentially simulates the graded subset selection of the red objects, and the commensurate dampening of the nonred objects, that may be instigated when the participant first hears "red" while the display is visible (and searchable) and then hears "vertical."

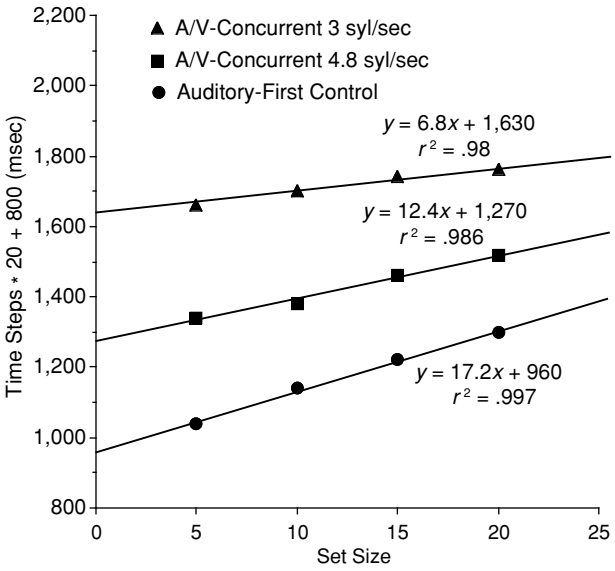

Figure 8. Simulation of Experiment 1 (and of Gibson et al., 2005) with the normalized recurrence competition algorithm. Greater temporal separation between the recognition point of the first adjective and that of the second adjective (with slower speech) allows for more effective suppression of the objects that do not exhibit the first adjective's feature. The result is a more efficient search (shallower slope) across different set sizes.

With faster speech, such as that used in Gibson et al. (2005), there is less time for the redness vector's cross talk to exert this "spreading suppression" (see Duncan & Humphreys, 1989) on the nonred objects in the verticalness vector. Thus, when the verticalness vector receives input, because the word *vertical* is recognized, the nodes corresponding to nonred vertical objects have been reduced by only a fifth of their starting activation, from .01 to .0079. Thus, their input causes them to compete just enough to impede the efficiency with which the salience of the target can emerge from that of the distractors.

Since the normalization forces all activations within any vector in this network to share a resource of 1.0 salience, this model has some properties in common with limited-capacity parallel models of search (see, e.g., Townsend & Ashby, 1983). However, the temporal dynamics of its activations are smoothly nonlinear. The normalized recurrence competition algorithm routinely produces sigmoidal activation curves for the winning representation that eventually takes up the probability space (Spivey &

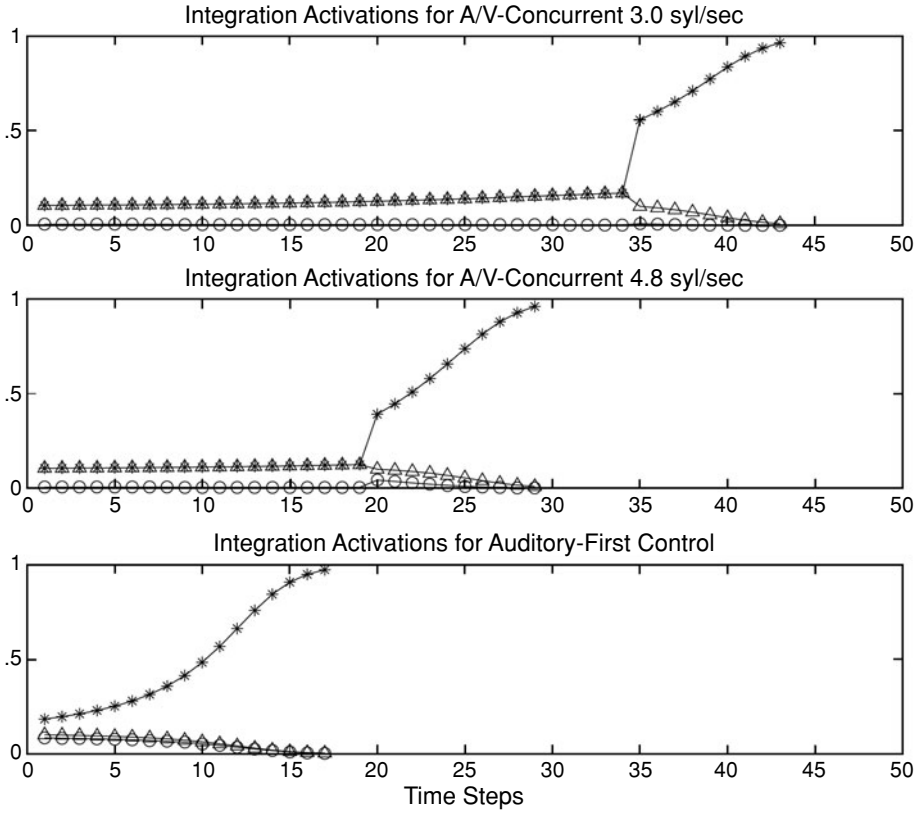

Figure 9. Activation curves from the normalized recurrence simulation of "Is there a red vertical?" with slower concurrent speech (upper panel), with faster concurrent speech (middle panel), and in the auditory-first control condition (lower panel)—each with a set size of 10. Note that in the upper and middle panels, the integration nodes corresponding to the red vertical bar (asterisks) and the red horizontal bars (triangles) slowly rise over time, but when the word *vertical* is recognized they sharply diverge. In the slower speech condition, the nodes corresponding to the green vertical bars (circles) are too suppressed to rise much at all when the "vertical" input is received. In contrast, in the faster speech condition, they rise just enough that their competition prevents the rise of the red vertical node from being quite as sharp as it is in the slower speech condition.

Dale, 2004). Figure 9 shows the activations of the integration vector for the three simulations in Figure 8. The model clearly suggests that the improved efficiency in A/V-concurrent visual search (top panel) lies predominantly in the processing of the last adjective—the one that singles out a unique object within the salience-enhanced subset. Within a probabilistic competitive vector, a single uniquely high activation will take over the probability space quite quickly.

Experiment 2 was simulated through the use of differential weights for the two feature vectors: a greater weight for the redness vector (2/3) and a lesser weight for the verticalness vector (1/3), as was suggested in the Discussion section of that experiment. The model's behavior is sufficiently robust that this weight change did not alter the results for the color-then-orientation simulation at all (cf. Figures 8 and 10). However, the orientation-then-color simulation produced a slightly less efficient search function, as was observed in Experiment 2 (see Figure 10). In this orientation-then-color simulation, the longer time between the recognition point of the three-syllable adjective *vertical* and that of the one-syllable adjective *red* (45 time steps instead of 35) would normally allow for *greater* suppression of the nonvertical objects in the redness vector, but the weaker weight on the verticalness vector resulted in overall *weaker* suppression. Therefore, the orientation-then-color condition produced a slightly less efficient RT $\times$ set size slope than did the color-then-orientation condition. (When these new weights are used to repeat the first set of simulations, results are almost identical to those in Figure 8.)

In our final simulation, of Experiment 3, three feature vectors were used to simulate results for the query "Is there a red vertical large?" For the A/V-concurrent condition, 35 time steps after the redness vector was given its input, the verticalness vector was given its input; 45 time steps after that, the largeness vector was given its input. With the redness vector continuing to have a weight twice that of the verticalness vector (2/5 vs. 1/5) and equal to that of the largeness vector (2/5), and with all other parameters kept the same as in the previous simulations, the results mimic the data from Experiment 3 reasonably well (see Figure 11). The slope for the auditory-first condition

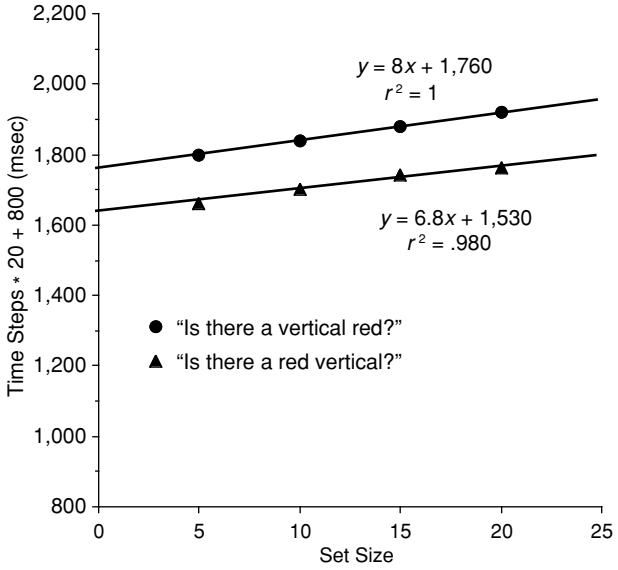

Figure 10. With a greater weight (or overall salience) for the redness vector and a lesser weight for the verticalness vector, this simulation produces the subtle asymmetry between adjective orders observed in the A/V-concurrent conditions of Experiment 2.

fits that of the human data almost perfectly. However, the RTs are a little short overall, and the slope for the A/V-concurrent triple-conjunction simulation is a few milliseconds per item shallower than that of the human data.

These simulations do not prove that our linguistic assistance of visual search is conducted via a biased competition framework (Desimone & Duncan, 1995) that interacts fluidly with language processes, in which the graded salience of object representations is updated in parallel as each adjective is heard. Nonetheless, they serve as useful existence proofs that such a framework could indeed account for these results. As each adjective is heard, it may trigger the graded enhancement of all objects that exhibit the corresponding feature. As these salience-enhanced objects compete, they suppress the salience of the objects that do not exhibit that adjective's feature. When the last adjective is heard in these A/V-concurrent conditions, it would be able to enhance the salience of all objects that exhibit its feature, but the only such object that has not already been suppressed in salience is the target object. This uniquely salient object representation can then win the biased competition quickly and efficiently.

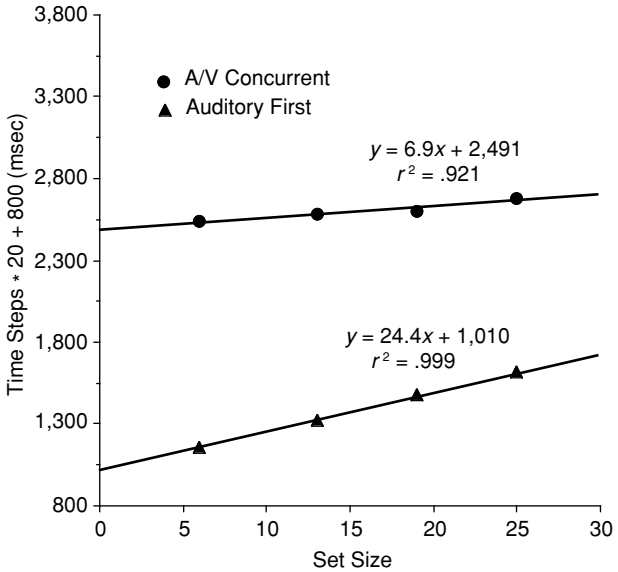

Figure 11. With three feature vectors, this simulation approximates the results of Experiment 3. However, it predicts a slightly shallower slope for the A/V-concurrent condition than was obtained in the human data.

The present combination of perceptual experiments and quantitative simulations extends the generalizability of the phenomenon of linguistically mediated visual search to a variety of experimental circumstances, rules out some alternative explanations, and points toward some potential mechanisms that could explain it. Future research on this phenomenon will benefit greatly from combining human experimentation with the further development of explicit computational models of this kind of interaction between language and vision. As implemented models of spoken word recognition are interfaced with implemented models of visual processing (see, e.g., Roy & Mukherjee, 2005; Spivey, Grosjean, & Knoblich, 2005), we can begin to formulate a richer understanding of exactly how language comprehension and visual perception manage to interact so fluidly.

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 & Language*, **38**, 419-439.

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

Arnfield, S., Roach, P., Setter, J., Greasley, P., & Horton, D. (1995). Emotional stress and speech tempo variation. In I. Trancoso & R. Moore (Eds.), *Proceedings of the ESCA–NATO Tutorial and Research Workshop on Speech Under Stress* (pp. 13-15). Lisbon: ISCA Archive.

Boucart, M., & Humphreys, G. W. (1997). Integration of physical and semantic information in object processing. *Perception*, **26**, 1197-1209.

Cavanagh, P. (1987). Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity and shape. *Computer Vision, Graphics, & Image Processing*, **37**, 171-195.

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

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, **30**, 39-78.

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

Donk, M., & Theeuwes, J. (2001). Visual marking beside the mark: Prioritizing selection by abrupt onsets. *Perception & Psychophysics*, **63**, 891-900.

Donk, M., & Verburg, R. C. (2004). Prioritizing new elements with a

brief preview period: Evidence against visual marking. *Psychonomic Bulletin & Review*, **11**, 282-288.
Dosher, B. A., Han, S., & Lu, Z. L. (2004). Time course of asymmetric visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **30**, 3-27.
Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**, 433-458.
Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, **24**, 409-436.
Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*, **9**, 111-118.
Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 32-39.
Gibson, B. S., Eberhard, K. M., & Bryant, T. A. (2005). Linguistically mediated visual search: The critical role of speech rate. *Psychonomic Bulletin & Review*, **12**, 276-281.
Haslam, N., Porter, M., & Rothschild, L. (2001). Visual search: Efficiency continuum or distinct processes? *Psychonomic Bulletin & Review*, **8**, 742-746.
Horowitz, T. S., & Wolfe, J. M. (2003). Memory for rejected distractors in visual search? *Visual Cognition*, **10**, 257-298.
Kaptein, N. A., Theeuwes, J., & van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1053-1069.
Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, **240**, 740-749.
Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The microstructure of spoken word recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, **132**, 202-227.
McClelland, J. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, **86**, 287-330.
McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 1517-1537.
Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, **57**, 1-17.
Nakayama, K., & Joseph, J. S. (1998). Attention, pattern recognition, and pop-out in visual search. In R. Parasuraman (Ed.), *The attentive brain* (pp. 279-298). Cambridge, MA: MIT Press.
Olds, E. S., Cowan, W. B., & Jolicœur, P. (2000a). Partial orientation pop-out helps difficult search for orientation. *Perception & Psychophysics*, **62**, 1341-1347.
Olds, E. S., Cowan, W. B., & Jolicœur, P. (2000b). The time-course of pop-out search. *Visual Research*, **40**, 891-912.
Olds, E. S., & Fockler, K. A. (2004). Does previewing one stimulus feature help conjunction search? *Perception*, **33**, 195-216.
Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, **40**, 1227-1268.
Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, **109**, 160-174.
Quinlan, P. T., & Humphreys, G. W. (1987). Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. *Perception & Psychophysics*, **41**, 455-472.
Reynolds, J. H., & Desimone, R. (2001). Neural mechanisms of attentional selection. In J. Braun & C. Koch (Eds.), *Visual attention and cortical circuits* (pp. 121-135). Cambridge, MA: MIT Press.
Roy, D., & Mukherjee, N. (2005). Toward situated speech understanding: Visual context priming of language models. *Computer Speech & Language*, **19**, 227-248.
Sagi, D., & Julesz, B. (1984). Detection versus discrimination of visual orientation. *Perception*, **13**, 619-628.
Spivey, M. J., & Dale, R. (2004). On the continuity of mind: Toward a dynamical account of cognition. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 45, pp. 87-142). San Diego: Elsevier, Academic Press.
Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, **102**, 10393-10398.
Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, **45**, 447-481.
Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, **12**, 282-286.
Spivey-Knowlton, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (1998). Integration of visuospatial and linguistic information in real time and real space. In P. Olivier & K.-P. Gapp (Eds.), *Representation and processing of spatial expressions* (pp. 201-214). Mahwah, NJ: Erlbaum.
Spratling, M. W., & Johnson, M. H. (2004). A feedback model of visual attention. *Journal of Cognitive Neuroscience*, **16**, 219-237.
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.
Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, **40A**, 201-237.
Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, **12**, 97-136.
Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 459-478.
Watson, D. G., & Humphreys, G. W. (1997). Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review*, **104**, 90-122.
Watson, D. G., & Humphreys, G. W. (2002). Visual marking and visual change. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 379-395.
Wolfe, J. M. (1992). "Effortless" texture segmentation and "parallel" visual search are *not* the same thing. *Vision Research*, **32**, 757-763.
Wolfe, J. M. (1994). Guided Search 2.0: A revised mode of visual search. *Psychonomic Bulletin & Review*, **1**, 202-238.
Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, **9**, 33-39.
Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 483-502.
Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 419-433.
Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 121-138.

NOTE

1. It is worth noting that average speaking rates in English range from 3 to 6 syllables/sec, depending on the speaker and the circumstances (Arnfield, Roach, Setter, Greasley, & Horton, 1995). Therefore, the speaking rate used in Spivey et al. (2001) and in the present experiments (3 syllables/sec) is on the slower end of the continuum but still within the normal range.

(Manuscript received July 8, 2004;
revision accepted for publication September 14, 2005.)