Research Article

Attentional Cues in Real Scenes, Saccadic Targeting, and Bayesian Priors

Miguel P. Eckstein, Barbara A. Drescher, and Steven S. Shimozaki

University of California, Santa Barbara

ABSTRACT—*Performance finding a target improves when artificial cues direct covert attention to the target's probable location or locations, but how do predictive cues help observers search for objects in real scenes? Controlling for target detectability and retinal eccentricity, we recorded observers' first saccades during search for objects that appeared in expected and unexpected locations within real scenes. As has been found with synthetic images and cues, accuracy of first saccades was significantly higher when the target appeared at an expected location rather than an unexpected location. Observers' saccades with target-absent images make it possible to distinguish two mechanisms that might mediate this effect: limited attentional resources versus differential weighting of information (Bayesian priors). Endpoints of first saccades in target-absent images were significantly closer to the expected than the unexpected locations, a result consistent with the differential-weighting model and inconsistent with limited resources being the sole mechanism underlying the effect.*

Artificial cues such as boxes or arrows indicating the possible location of a target often improve human search performance in a variety of perceptual tasks, such as target detection (Palmer, Ames, & Lindsey, 1993; Palmer, Verghese, & Pavel, 2000), identification (Bennett & Jaye, 1995; Talgar, Pelli, & Carrasco, 2004), localization (Burgess & Ghandeharian, 1984), and orientation discrimination (Baldassi & Verghese, 2002; Morgan, Ward, & Castet, 1998). Furthermore, when a cue's validity is probabilistic (e.g., 80%), perceptual performance and accuracy of the first search saccade are typically better in the trials in which the target appears at the cued location (valid-cue trials) than in the trials in which it appears at the uncued location (invalid-cue trials; Bashinski & Bacharach, 1980; Downing, 1988; Posner, 1980; Shimozaki, Eckstein, & Abbey, 2003; Shimozaki, Schoonveld, & Eckstein, 2005). These performance improvements with cue validity (i.e., cuing effect) have been attributed to shifts of visual attention to the cued locations.

How might the mechanisms that benefit perceptual-saccadic decisions with such synthetic cues and images improve performance when humans search for objects in real scenes? Real scenes do not have high-contrast arrows or boxes indicating the probable locations of objects that might be searched for. Yet objects do not appear at random locations in scenes, but often co-occur with other objects or with visual properties of the background (e.g., a specific color) that might be highly visible. For example, if observers are instructed to search for the chimney in a scene of a house (see Fig. 1, top), their first-saccade accuracy might be poor if the target is not highly visible in the visual periphery (Beutter, Eckstein, & Stone, 2003; Findlay, 1997; Zelinsky & Sheinberg, 1997). However, they might use a highly visible object or visual properties of the background that often co-occur with the low-visibility target as a cue to its location (Chun, 2000), much in the way they use the boxes or arrows that cue the target location in the synthetic images created in the laboratory. This strategy would be useful on most occasions. However, it would not be beneficial in the unlikely scenarios in which the chimney appears at an unexpected location (see Fig. 1, middle). Thus, if highly visible objects act as cues for target locations in real scenes, one would expect the typical cuing effect found with synthetic cues and images (Posner, 1980; Shimozaki et al., 2003) to generalize to search in real scenes: The accuracy of finding a target when it appears with a high-visibility object or visual property with which it often co-occurs in scenes (expected location $\approx$ valid cue) should be higher than the accuracy of finding that same target when it

Address correspondence to Miguel P. Eckstein, Department of Psychology, University of California, Santa Barbara, Santa Barbara, CA 93106, e-mail: eckstein@psych.ucsb.edu.

Fig. 1. Examples of a real scene with a target at an expected location (chimney on house roof), the target at an unexpected location (chimney next to tree), and the target absent (no chimney). The squares in the images correspond to the endpoints of observers' first saccades.

appears with an object or visual property with which it rarely co-occurs (unexpected location $\approx$ invalid cue).

A number of previous studies have found improved accuracy of perceptual judgments and search saccades when target objects appear in semantically consistent images or at expected locations (Biederman, 1981; Henderson, 2003; Henderson & Hollingworth, 1998; Henderson, Weeks, & Hollingworth, 1999). These facilitation effects have been referred to as context effects, and researchers have related such effects to the deployment of visual attention (Chun, 2000; Oliva, Torralba, Castelhano, & Henderson, 2003). However, no study of human search for targets that are semantically consistent with the scenes in which they appear has directly compared the accuracy of first saccades for targets that appear at expected versus unexpected locations within the scenes. In the study reported here, we measured observers' first saccades during search for objects located at expected and unexpected locations to determine whether there is a target-location effect on saccade accuracy in real scenes analogous to the cuing effects observed with synthetic cues. To discount low-level visual explanations of the context effect, we controlled for retinal eccentricity and included a control condition in which local backgrounds were cropped (to verify that target detectability was, on average, constant across the expected- and unexpected-location conditions).

The mechanism responsible for context effects in real scenes remains an important unanswered question (Chun, 2000). The classic cuing effect with perceptual judgments and synthetic cues and images has been explained in terms of two competing theories. The standard explanation is that the effect is due to the allocation of limited attentional resources (Bashinski & Bacharach, 1980; Luck, Hillyard, Mouloua, & Hawkins, 1996). In this classic model, the cue allows the observer to allocate attentional resources to a single location, rather than distribute them across many locations, and therefore enhances the quality of processing at that cued (attended) location relative to processing at uncued locations (e.g., Downing, 1988; Hawkins et al., 1990). The alternative explanation of the cuing effect is that it is due to differential weighting of visual information arising from the cued and uncued locations. In this model, perceptual and saccadic decision errors arise because distractors are mistakenly confused with the target as a result of noise in the brain's internal response to each display element (Green & Swets, 1966; Palmer et al., 1993; Shaw, 1982). In this framework, attention allows the observer to weight information arising from the cued location, which is likely to contain the target, more heavily than information arising from the uncued location, which is likely to contain only noise (Eckstein, Shimozaki, & Abbey, 2002; Kinchla, Chen, & Evert, 1995; Shimozaki et al., 2003; Fig. 2). This strategy will maximize performance across all types of trials and, as a by-product, give rise to a cuing effect; thus, in this model, there is no need for researchers to postulate limited attentional resources or enhancement of processing at the cued location (i.e., sensitivity change). The theory of Bayesian decision making specifies the mathematically optimal method (i.e., the method that maximizes performance across all trials) of weighting noisy visual information from the cued and uncued locations on the basis of prior knowledge about the validity of the cues (Eckstein et al., 2002; Kersten, Mamassian, & Yuille, 2004; Shimozaki et al., 2003). The weight for each location is often referred to as a Bayesian prior.

The main goal of this study was to identify whether the mechanism mediating context effects in saccadic targeting is limited attentional resources or differential weighting of information. Greater accuracy of observers' first saccades for images with the target at the expected location than for images with the target at an unexpected location would be consistent with both models of attention. However, a third critical set of images—images of the scenes without the targets (target-absent condition; Fig. 1, bottom)—allowed us to distinguish the two models. In the differential-weighting model, evidence from the expected location is weighted more highly than evidence from unexpected locations, so the model is prone to choose to make an eye movement to the expected location rather than other locations, even when the target is absent. In contrast, in the limited-resources model, the resource allocation to the expected target location leads to improved perceptual processing of that location, and therefore both better ability to detect the target when it is present at the expected location and better ability to reject the expected location when the target is absent. Thus, when the target is absent from the image, the limited-resources model, unlike the differential-weighting model, will not be prone to make an eye movement toward the expected location. Therefore, measuring observers' first saccades for images in which the target was absent allowed us to determine whether the context effect was consistent with one or the other model.

Next we describe mathematical versions of the two models, as applied to the tasks used in the present study. We include this information to emphasize that it is possible to implement testable computational models that are consistent with the concepts and predictions just outlined. The predictions of the model simulations are presented in the Results section, and mathematical details are given in the appendix.

COMPUTATIONAL MODELS

Differential-Weighting Model (Bayesian Priors)

This model assumes that each of the considered locations in the scene elicits an internal response for each target-relevant visual feature. All internal responses are subject to Gaussian independent neural noise. The model then calculates for each location a joint likelihood of observing the feature responses given that the target is present at that location and a joint likelihood of observing the feature responses given that the target is absent at that location. The ratios of these two likelihoods distributed over the scene create a likelihood-ratio map (a map of evidence for target presence). The evidence is weighted by a map of weights specifying the prior probabilities of the target appearing at each location. The map of weights, or priors map, is based on the probability that other highly visible objects and visual properties in the background co-occur with the searched-for target object. The product of the likelihood-ratio and priors maps gives rise to a map of posterior probabilities of the target being present at each location. The first saccade is directed toward the location with maximum activity in the posterior-probability map.

Figure 2 shows a general schematic of the differential-weighting model for the current task and also for the cuing paradigm with synthetic cues. The model predicts that the accuracy of the first saccade will be greater when the target is present at an expected location than when it is present at an unexpected location because visual evidence for the presence of the target is weighted more heavily at the expected location. The implemented model has a number of additional components that have been included for realism but do not alter the dissociation in predictions across the two models tested: (a) uncertainty about the feature contrasts, requiring that likelihoods be summed across many possible feature contrasts; (b) a cost function penalizing large saccades over short saccades (Araujo, Kowler, & Pavel, 2001); and (c) a spatial Gaussian-distributed motor-error noise that perturbs the location targeted by the first saccade.

Limited-Attentional-Resources Model

In this model, attentional resources are deployed at likely target locations, which are cued by other highly visible objects. The particular computational model implemented in this study assumes noisy processing of features and includes an ideal decision rule following the limited-resource processing. The model starts with the extracted feature responses, which are subject to noise. Feature responses at the location or locations where attentional resources are deployed (expected locations) are processed with lower noise variability than feature responses at other locations, allowing for improved quality of processing. A likelihood ratio is calculated for each location, taking into account the known variance of the responses to expected and unexpected target locations. The model generates a saccade to the location with the highest likelihood ratio. Accuracy will be better when the target is at the expected location rather than at an unexpected location because the lower noise for the expected location (due to the allocation of attentional resources) will on average result in a likelihood ratio that is larger (more evidence of target presence) than the likelihood ratios for other locations (see the appendix for details).[1]

[1]Other implementations of the limited-resources model utilizing components that lead to an increased difference in mean response to target versus nontarget at the expected location or to a temporal serial processing of items in the display will result in qualitative predictions identical to those of the limited-resources model implemented in the current study, which used a varying internal noise at expected and unexpected locations. A saccade cost function that penalizes large saccades and observer uncertainty about the strength of the response to the target feature are included in the implemented models for the sake of completeness and do not change the predictions of the differential-weighting and limited-resources models for the target-absent images.

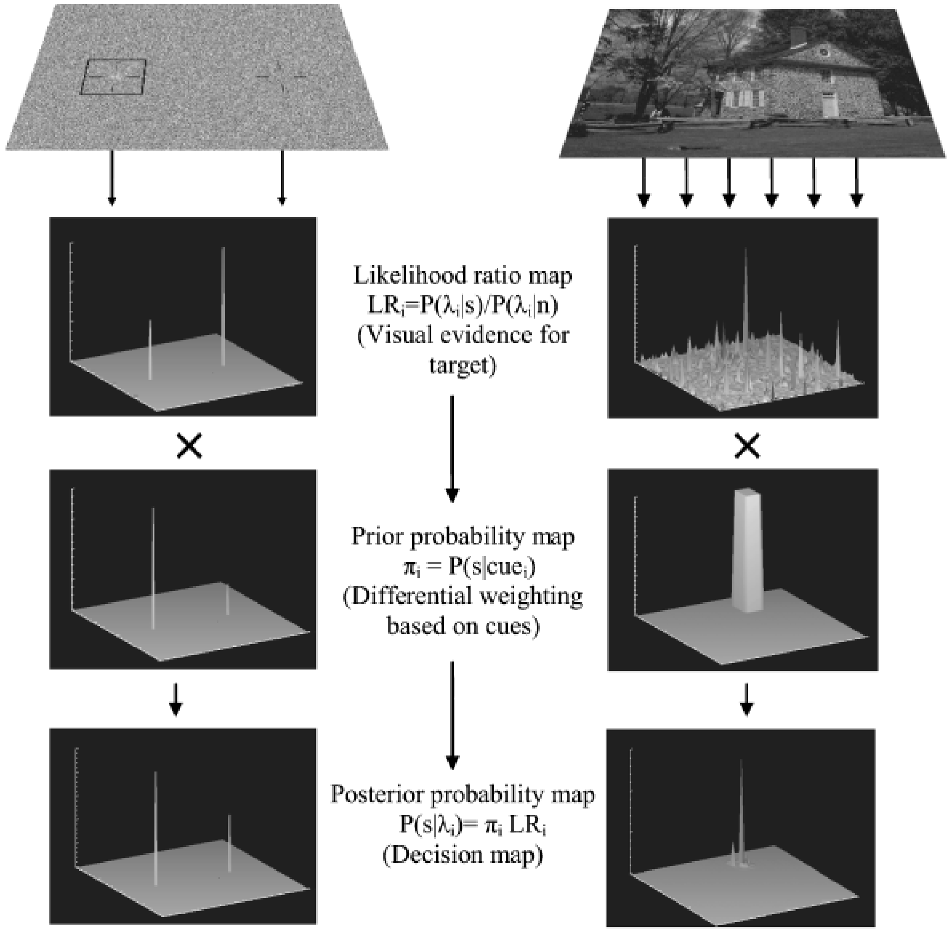

Fig. 2. Bayesian framework for weighting the evidence of target presence (likelihood ratio, LR_i, given the observed model response at the *i*th location, λ_i) at each location by the prior probability (π_i) of the target co-occurring with a highly visible cue. The illustrations on the left show how a decision map might be derived in a laboratory task in which the target is a bright blob that might appear in one of two locations. The cue is the black box shown in the top illustration; it indicates that if the target appears, it will appear at that location 80% of the time. The illustrations on the right show how this Bayesian framework may be generalized to search for objects in natural scenes. In this case, the target is the chimney, and the cues are highly visible objects (e.g., a house) that typically co-occur with chimneys. $P(\lambda_i|s)$ denotes the probability of observing the response λ_i given the presence of the target, and $P(\lambda_i|n)$ is the probability of observing the response λ_i given the absence of the target. $P(s|cue_i)$ is the probability of the target co-occurring with a cue detected at the *i*th location, and $P(s|\lambda_i)$ is the posterior probability of the presence of the target at the *i*th location given the observed response, λ_i. For simplicity, the schematic does not show additional components of the implemented Bayesian computational model (see the appendix).

METHOD

The study included three experimental conditions: (a) target in an expected location, (b) target in an unexpected location, and (c) target absent. All stimuli were real scenes, and the same scenes and objects were used for all three conditions. Objects (mean size = $3.63° \times 2.22°$) were placed in semantically consistent scenes ($36.5° \times 27.8°$) using Paint Shop Pro. Scenes with targets at the expected location included one showing a lamp on a bedside table and another showing keys on a living-room table. Scenes with targets at an unexpected location included one showing a lamp on a bed and another showing keys on the floor. To ensure equal retinal eccentricity across the expected- and unexpected-location conditions, we presented a fixation cross equidistant from the two possible target locations before presenting the real scene (average = 13.75°).

Twenty naive observers participated in 24 trials in which 24 different real scenes were presented. Random assignment of the participants to three groups ensured that each participant viewed each background once and only once. One third of the images assigned to each group contained target objects in the expected location, one third contained target objects in an unexpected location, and one third were target-absent images. Trials began with a 1-s fixation, followed by a word that named the target object, presented for 2 s. Following another fixation, the image appeared for 2 s. After the offset of the image, observers clicked a computer mouse to indicate if the object had appeared in it ("yes" or "no"). Images appeared in random order. Observers were not informed about the proportion of images with target present or absent.

Eye position was monitored using an infrared video-based eyetracker sampled at 250 Hz (Eyelink I, SMI/SR Research, Berlin, Germany). A chin rest was used to minimize head movements. A calibration was performed prior to initiation of the study. An eye movement was recorded as a saccade if both velocity and acceleration exceeded a threshold (velocity $> 35°/s$; acceleration $> 9500°/s^2$). The first saccade outside an area extending 2.5° from the initial fixation was considered the first goal-oriented saccade.

We ran a control condition with an additional 13 observers to ensure that the detectability of the targets against the different backgrounds did not differ significantly between the expected- and unexpected-location conditions. Observers viewed two cropped circular areas (diameter = 5.6°) from each scene (presented 6.3° to the left and right of central fixation), one containing the object on the background and the other containing the background only. The cropping removed the object's context. The positions (left vs. right) of the cropped images were randomized. The participants' task was to identify the image that contained the target (two-alternative forced choice). Trials started with a central fixation, followed by presentation of a word naming the target object, another central fixation, and the cropped images, which were displayed for 50 ms.

RESULTS

Figure 1 (top and middle scenes) shows the endpoints of the first saccades for all observers for a representative image with the target (chimney) at the expected and unexpected locations. Figure 3 (upper left graph) shows the mean distance from the endpoint of the first saccade to the target location, averaged across images and observers for both conditions. Distance to the target was significantly greater in the unexpected-location condition than in the expected-location condition, $t(19) = -5.25$, $p_{rep} > .999$,[2] $\eta^2 = .59$. Figure 3 (middle and lower left graphs) also shows that both the limited-resources model and the differential-weighting model predicted increased saccadic accuracy when the target was at an expected location.

When no target object was present, observers' first search saccades were significantly closer to the expected target location than to the unexpected location (Fig. 1, bottom scene; Fig. 3, upper right graph), $t(19) = -6.68$, $p_{rep} > .999$, $\eta^2 = .70$. Figure 3 also shows that the differential-weighting model predicted that the first saccade's endpoint would be closer to the expected target location than to the unexpected target location (middle right graph), whereas the limited-resources model predicted no significant difference in the distance to the two locations (lower right graph).

The mean latency of the first saccade was not significantly different across the conditions (236 ms for the expected-location condition, 225 ms for the unexpected-location condition, and 228 ms for the target-absent condition), $F(2, 38) < 1$.

Accuracy in the two-alternative forced-choice target-detection task with cropped backgrounds did not differ significantly between the images cropped from the expected target location (75.8% correct) and those cropped from the unexpected target location (77.1%), $t(12) < 1$.

[2]For an interpretation of p_{rep} and information on how it is calculated, see Killeen (2005).

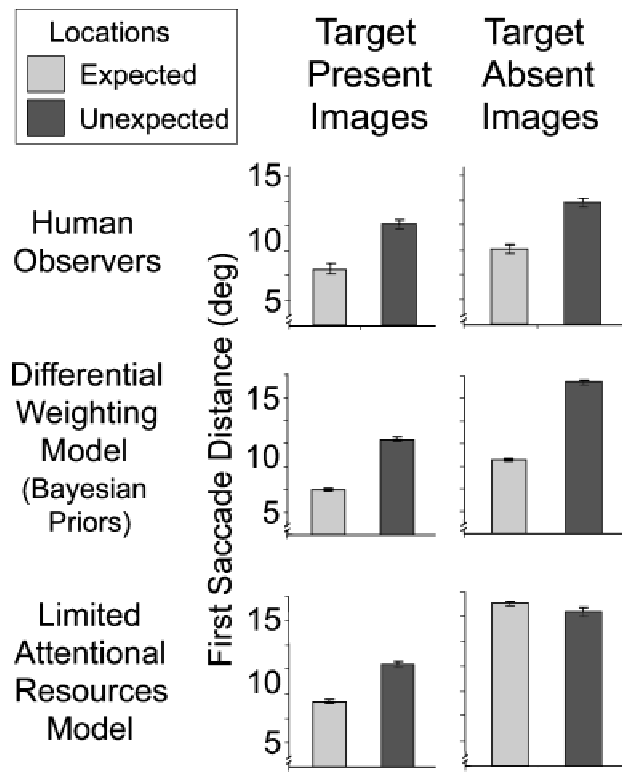

Fig. 3. Experimental results (top row) and predictions of the differential-weighting (middle row) and limited-resources (bottom row) models. The graphs in the left column show the mean distance of the first saccade's endpoint from the target location for images with the target at the expected location and images with the target at an unexpected location. The graphs in the right column show the mean distance of the first saccade's endpoint from the expected and unexpected locations in target-absent images. The model predictions are based on computer simulations.

DISCUSSION

The results show that the accuracy of the first saccade was significantly higher when the target appeared at an expected location rather than an unexpected location. This effect cannot be explained in terms of low-level visual variables (target detectability or retinal eccentricity). Thus, we conclude that the increase in accuracy of the first saccade when the target is at an expected location can be attributed to observers' expectations about where the target will be. According to this explanation, some portion of the scene's background or a highly visible object that often co-occurs with the target object acts as a cue to the target's location. For an object or background to serve as a cue requires a fast recognition system that can identify parts of the scene within 200 ms, and in fact such recognition speed has been reported recently in a number of studies (e.g., VanRullen & Thorpe, 2001). The present results also agree with recent studies showing that context and environmental structure guide the deployment of saccades during search of real scenes (Hayhoe & Ballard, 2005; Oliva et al., 2003).

A central issue in the field has been to clarify the mechanism by which attention mediates these context effects (Chun, 2000). In our study, when the target was absent, observers' first saccades were significantly closer to the expected target location than to the unexpected location. These results are consistent with differential weighting of information (Eckstein et al., 2002; Torralba, 2003), and not with limited attentional resources being the sole mechanism mediating context effects. Although we implemented differential weighting in terms of a Bayesian observer, other weighting models, including linear (Kinchla et al., 1995) and connectionist models (McClelland & Rumelhart, 1981), might also account for the results. The results should also not be interpreted as implying that priors for a given target object and kind of scene are compulsorily used by an observer to search through every single scene of that type. Thus, although most people would place a telephone on a living-room table rather than under the couch, you may look for your phone under the couch in your own living room if, in fact, that is a place where you often put your phone (Henderson, 2003). Thus, priors can be scene-specific.

Together, our results provide some insight into important components that must be included in computational models of saccadic targeting. Currently, some models do not incorporate knowledge about searched-for targets (Itti & Koch, 2000), and others do (for noise images, see Beutter et al., 2003; for real images, see Rao, Zelinsky, Hayhoe, & Ballard, 2002). In addition, a recently proposed model includes knowledge of varying target visibility across the retina (Najemnik & Geisler, 2005). The present results suggest that computational models of saccadic targeting should also include information about context, much in the same way models of attention use priors to weight information differentially on the basis of cues (Eckstein et al., 2002; Shimozaki et al., 2003, 2005). Indeed, Oliva et al. (2003) and Torralba (2003) have recently developed a model that extracts from scenes statistical information about low-level features that typically co-occur with the target and uses this information as priors in decisions about target location and recognition.

To summarize, the present experiment suggests that context effects on human saccadic targeting are consistent with the concept of differential weighting of information and inconsistent with the idea that limited resources are the sole attentional mechanism responsible for the observed effects.[3] Finally, our results show that two distinct behavioral effects (cuing effects in synthetic images and context effects in real scenes) might be explained with a common mechanism: differential weighting of visual information on the basis of co-occurrence of targets and cues.

Acknowledgments—This research was supported by National Science Foundation Grant 0135118.

REFERENCES

Araujo, C., Kowler, E., & Pavel, M. (2001). Eye movements during visual search: The costs of choosing the optimal path. *Vision Research*, *41*, 3613–3625.

Baldassi, S., & Verghese, P. (2002). Comparing integration rules in visual search. *Journal of Vision*, *2*, 559–570.

Bashinski, H.S., & Bacharach, V.R. (1980). Enhancements of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception & Psychophysics*, *28*, 241–248.

Bennett, P.J., & Jaye, P.D. (1995). Letter localization, not discrimination, is constrained by attention. *Canadian Journal of Experimental Psychology*, *49*, 460–504.

Beutter, B.R., Eckstein, M.P., & Stone, L.S. (2003). Saccadic and perceptual performance in visual search tasks. I. Contrast detection and discrimination. *Journal of the Optical Society of America A*, *20*, 1341–1355.

Biederman, I. (1981). Do background depth gradients facilitate object identification? *Perception*, *10*, 573–578.

Burgess, A.E., & Ghandeharian, H. (1984). Visual signal detection. II. Signal-location identification. *Journal of the Optical Society of America A*, *8*, 906–910.

Chun, M.M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*, 170–178.

Downing, C.J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 188–202.

Eckstein, M.P., Shimozaki, S.S., & Abbey, C.K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *Journal of Vision*, *2*, 25–45.

Findlay, J.M. (1997). Saccade target selection during visual search. *Vision Research*, *37*, 617–631.

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hawkins, H.L., Hillyard, S.A., Luck, S.J., Mouloua, M., Downing, C.J., & Woodward, D.P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 802–811.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*, 188–194.

Henderson, J.M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498–504.

Henderson, J.M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance while reading and while watching dynamic scenes* (pp. 269–293). Oxford, England: Elsevier.

[3]Note that a mixture model with both differential weighting and a relatively small degree of resource limitation might also be consistent with the findings.

Henderson, J.M., Weeks, P.A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 210–228.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Killeen, P.R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, *16*, 345–353.

Kinchla, R.A., Chen, Z., & Evert, D. (1995). Precue effects in visual search: Data or resource limited? *Perception & Psychophysics*, *57*, 441–450.

Luck, S.J., Hillyard, S.A., Mouloua, M., & Hawkins, H.L. (1996). Mechanisms of visual-spatial attention: Resource allocation or uncertainty reduction? *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 725–737.

McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

Morgan, M.J., Ward, R.M., & Castet, E. (1998). Visual search for a tilted target: Tests of spatial uncertainty models. *Quarterly Journal of Experimental Psychology A*, *51*, 347–370.

Najemnik, J., & Geisler, W.S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*, 387–391.

Oliva, A., Torralba, A., Castelhano, M.S., & Henderson, J.M. (2003). Top-down control of visual attention in object detection. In *Proceedings of the IEEE International Conference on Image Processing* (Vol. I, pp. 253–256). Piscataway, NJ: IEEE.

Palmer, J., Ames, C.T., & Lindsey, D.T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 108–130.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*, 1227–1268.

Peterson, W.W., Birdsall, T.G., & Fox, W.C. (1954). The theory of signal detectability. *Transactions of the Institute of Radio Engineers Professional Group on Information Theory*, *4*, 171–212.

Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.

Rao, R.P., Zelinsky, G.J., Hayhoe, M.M., & Ballard, D.H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*, 1447–1463.

Shaw, M.L. (1982). Attending to multiple sources of information: 1. The integration of information in decision-making. *Cognitive Psychology*, *14*, 353–409.

Shimozaki, S.S., Eckstein, M.P., & Abbey, C.K. (2003). Comparison of two weighted integration models for the cueing task: Linear and likelihood. *Journal of Vision*, *3*, 209–229.

Shimozaki, S.S., Schoonveld, W., & Eckstein, M.P. (2005, May). *An ideal observer approach to unifying set size and cueing effects for perceptual and saccadic decisions*. Paper presented at the annual meeting of the Vision Sciences Society, Sarasota, FL.

Talgar, C.P., Pelli, D.G., & Carrasco, M. (2004). Covert attention enhances letter identification without affecting channel tuning. *Journal of Vision*, *4*, 22–31.

Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America A*, *20*, 1407–1418.

VanRullen, R., & Thorpe, S.J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, *30*, 655–668.

Zelinsky, G.J., & Sheinberg, D.L. (1997). Eye movements during parallel-serial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 244–262.

(Received 8/1/05; Revision accepted 12/6/05; Final materials received 1/24/06)

APPENDIX

Differential-Weighting Model (Bayesian Priors)

The Bayesian model implemented in this study assumes the observer is monitoring J sensory units tuned to J different possible features of the target object. For each trial (t) and each considered image location (x,y), each sensory unit gives rise to a noisy response for each feature λ_j. The model calculates the likelihood of observing the response λ_j given the presence of the target's jth feature at that location and the likelihood of the response given the absence of that feature. Assuming that the sensory unit's response to the target's jth feature has a Gaussian distribution, a mean of d'_j and a standard deviation of σ, then the likelihood that the jth sensory unit takes a value $\lambda_{j,x,y,t}$ given the presence of the target's jth feature at location x,y on trial t is given by

$$l_{j,x,y,t}(\lambda_{j,x,y,t}|\mathrm{s}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \left[\frac{(\lambda_{j,x,y,t} - d'_j)^2}{2\sigma^2} \right], \qquad \text{(A1)}$$

where s stands for "signal" and denotes the presence of the target. Also, the likelihood that the jth sensory unit takes a value $\lambda_{j,x,y,t}$ given the absence of the target's jth feature is given by

$$l_{j,x,y,t}(\lambda_{j,x,y,t}|\mathrm{n}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \left[\frac{\lambda^2_{j,x,y,t}}{2\sigma^2} \right], \qquad \text{(A2)}$$

where n stands for "noise" and denotes the absence of the target.

A likelihood ratio, LR, can then be calculated (Green & Swets, 1966):

$$LR_{j,x,y,t} = \frac{l_{j,x,y,t}(\lambda_{j,x,y,t}|\mathrm{s})}{l_{j,x,y,t}(\lambda_{j,x,y,t}|\mathrm{n})} = \exp \left[\frac{\lambda_{j,x,y,t} d'_j - 0.5 d'^2_j}{\sigma^2} \right] \qquad \text{(A3)}$$

To combine information across feature dimensions optimally for each location, the observer multiplies the likelihoods across all features. Thus, a joint likelihood is computed for each location:

$$LR_{x,y,t} = \prod_{j=1}^{J} LR_{j,x,y,t} \qquad \text{(A4)}$$

Equations A1 through A4 assume that the observer knows a priori the strength of the internal response generated by the presence of the target. This is not a good assumption, given that in our experiment observers had not seen the targets prior to the search task. Thus, given the uncertainty about feature strength, the model calculates likelihood ratios for K different feature strengths (d'_k) and then the sum of the likelihood ratios (*SLR*; Peterson, Birdsall, & Fox, 1954):

$$SLR_{x,y,t} = \sum_{k=1}^{K} LR_{x,y,t,k} \quad \text{(A5)}$$

The model then computes a posterior probability that the target is present at each location (x,y) by multiplying the *SLR* at that location by the prior probability of the object being located at that location ($\pi_{x,y,t}$):

$$P_{x,y,t}(\mathrm{s}|\lambda) = \pi_{x,y,t} SLR_{x,y,t} \quad \text{(A6)}$$

The prior probability ($\pi_{x,y,t}$) is given by the probability of the presence of the target object given a highly visible background object or visual property.

The posterior probabilities are multiplied by a cost function (a Gaussian function, $C_{x,y,t}$, centered at the initial point of fixation for each image, with standard deviation sampled independently for each observer) that differentially penalizes large saccades over short saccades (Araujo et al., 2001). An eye movement for trial t, $\mathrm{EM}_{x,y}(t)$, is then generated to the location (x,y) for which the product of the posterior probability and the cost function is highest:

$$\mathbf{EM}_{x,y}(t) = Loc(Max(P_{x,y,t}(\mathrm{s}|\lambda)C_{x,y,t}) + \varepsilon_{x,y}(t)), \quad \text{(A7)}$$

The *Max* function takes the maximum value, among the considered locations, of the products of the posterior probabilities and cost values. The *Loc* function identifies the x,y coordinates of the location with the maximum value. $\mathbf{EM}_{x,y}(t)$ is a vector containing the x and y coordinates. The *Max* function takes the maximum value of the products of the posterior probabilities and cost values among the considered locations. The *Loc* function identifies the x,y coordinates of the location with the maximum value. A motor error consisting of two random variables sampled from a Gaussian distribution was added to the x and y locations: $\varepsilon_{x,y}(t) \sim N(0, \sigma_{em} = 0.378 \text{ deg})$.

Monte Carlo simulations were used to calculate the model's first-saccade accuracy. We used the target positions and fixations from the images used with the human observers. The mean sensor response for each of the target's features on each trial was randomly sampled from a uniform distribution (bounded by 1.0 and 4.0) with unit variance. The number of features was fixed to be three. The model calculated the likelihood of the response given that each of the three feature responses was sampled from 1 of 10 possible distributions, each with a different mean (means ranged from 1.0 to 5.5 in intervals of 0.5). The model considered x,y locations on a grid with neighboring points 0.236° apart. The priors for the expected location were randomly sampled from a uniform distribution bounded by 0.9 and 0.98 and assigned to 81 locations within an area 2.4° by 2.4°.

Limited-Attentional-Resources Model

The limited-resources model in this study was implemented within the framework of signal detection theory. The equations are identical to those presented for the Bayesian model with the exception that (a) the likelihood ratios (target evidence) are not weighted (i.e., not multiplied by priors) and (b) the variance of the internal response is inversely related to the product of the prior probability of a target being present at a given location and a constant (q). The lower internal noise for the expected target location relative to the unexpected locations makes the response less variable at the expected location and increases the likelihood ratio at the expected target location, where attentional resources have been deployed. In the limited-resources model, the likelihood ratio for each feature was calculated as follows:

$$\begin{aligned} LR_{x,y,t} &= \frac{l_{x,y,t}(\lambda|\mathrm{s})}{l_{x,y,t}(\lambda|\mathrm{n})} \\ &= \exp - \left[\frac{[(\lambda_j - d'_j)^2]}{2\sigma^2/\mathrm{q}\pi_{x,y,t}} - \frac{\lambda^2}{2\sigma^2/\mathrm{q}\pi_{x,y,t}} \right] \\ &= \exp \left[\frac{(\lambda d' - 0.5d'^2)}{2\sigma^2/\mathrm{q}\pi_{x,y,t}} \right] \end{aligned} \quad \text{(A8)}$$

After calculating a joint likelihood ratio of all three features, the model summed across likelihood ratios of possible feature strengths. The location with the maximum value of the product of the sum of the likelihood ratios and the cost function was chosen as the location to which the first saccade would be directed (with motor noise).