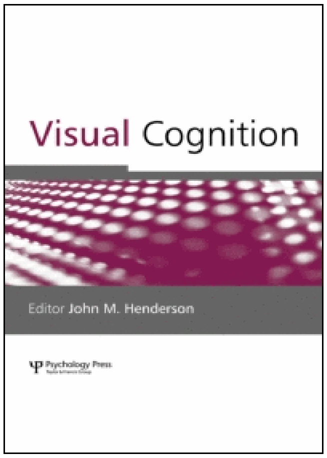

Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683696

Exploring set size effects in scenes: Identifying the objects of search

Mark B. Neider [a]; Gregory J. Zelinsky [a]
[a] Department of Psychology, Stony Brook University, Stony Brook, New York, USA

First Published on: 01 July 2007
To cite this Article: Neider, Mark B. and Zelinsky, Gregory J. (2007) 'Exploring set size effects in scenes: Identifying the objects of search', Visual Cognition, 16:1, 1 - 10
To link to this article: DOI: 10.1080/13506280701381691
URL: http://dx.doi.org/10.1080/13506280701381691

VISUAL COGNITION, 2008, 16 (1), 1–10

Psychology Press
Taylor & Francis Group

Exploring set size effects in scenes: Identifying the objects of search

Mark B. Neider and Gregory J. Zelinsky
Department of Psychology, Stony Brook University, Stony Brook, New York, USA

Traditional search paradigms utilize simple displays, allowing a precise determination of set size. However, objects in realistic scenes are largely uncountable, and typically visually and semantically complex. Can traditional conceptions of set size be applied to search in realistic scenes? Observers searched quasirealistic scenes for a tank target hidden among tree distractors varying in number and density. Search efficiency improved as trees were added to the display, a reverse set size effect. Eye movement analyses revealed that observers fixated individual trees when the set size was small, and the open regions between trees when the set size was large. Rather than a set size consisting of objectively countable objects, we interpret these data as evidence for a restricted functional set size consisting of idiosyncratically defined objects of search. Observers exploit low-level perceptual grouping processes and high-level semantic scene constraints to dynamically create objects that are appropriate to a given search task.

Despite its bedrock status in the search literature, the notion of a set size effect breaks down in the real world, where objects are difficult to delineate and typically outnumber the set sizes used in most search experiments. In these experiments, set size is defined as the number of items in a search array, with these items typically being simple patterns arranged on a homogeneous background (see Wolfe, 1998a, for a review). Search in the real world happens under very different conditions. Countless distractor objects may appear against a visually complex background, which itself may contain

Please address all correspondence to Gregory J. Zelinsky, Department of Psychology, State University of New York at Stony Brook, Stony Brook, NY 11794-2500, USA. E-mail: Gregory.Zelinsky@stonybrook.edu

Mark B. Neider is now at the Beckman Institute, University of Illinois at Urbana-Champaign. This work was supported by grants from the NSF (IIS-0527585), NIMH (R01-MH63748), and ARO (DAAD19-03-1-0039) to GJZ. The work described in this paper comprised part of the first author's doctoral dissertation submitted to Stony Brook University. We would like to thank Nancy Franklin for insightful comments on an earlier draft of this project, and Samantha Brotzen for her help with data collection.

http://www.psypress.com/viscog DOI: 10.1080/13506280701381691

object-like patterns (Neider & Zelinsky, 2006a). Moreover, grouping processes and semantic factors may constrain search to particular regions in real-world scenes (Neider & Zelinsky, 2006b; Torralba, Oliva, Castelhano, & Henderson, 2006; see Henderson & Hollingworth, 1999, for a review). These constraints create inequities among distractors; not all distractors are equally distracting. Not only is it difficult to count distractors in a realistic scene, it is therefore also difficult to know which distractors are relevant to the search task.

Generally, the search community has avoided the challenges posed by the conceptualization of set size in realistic contexts by separating the topics of set size from scene-based search. On the one hand, there is a rich tradition of characterizing search behaviour in terms of set size (e.g., Palmer, 1995; Wolfe, 1998b), but the stimuli in these studies were arrays of individuated objects, not scenes.[1] Consequently, they cannot address whether, or how, scene constraints affect the search set size function. On the other hand, studies exploring the scene-based factors affecting search have done so without appeal to the set size concept (e.g., Brockmole & Henderson, 2006; Henderson, Weeks, & Hollingworth, 1999; Neider & Zelinsky, 2006b; Toet, Kooi, Bijl, & Valeton, 1998; Torralba et al., 2006), leaving unaddressed any interaction between these factors and the number of objects in the scene. Recently, Rosenholtz, Li, Mansfield, and Jin (2005) suggested that a measure of visual clutter might be a reasonable surrogate for set size in realistic scenes. However, the relationship between clutter and set size is unclear. Introducing a textured background to a search display increases clutter, but does not affect the search slope (Neider & Zelinsky, 2006a; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002). Moreover, an objective measure of visual clutter would not be influenced by scene context, again leaving open the possible interaction between scene-based factors and set size. To date, there have been no systematic attempts to study search set size effects in a scene-based search task. The broad goal of this study is to provide an initial accounting of this relationship.

Our theoretical approach builds on the relevant set size paradigm popularized by Palmer (1995). In this paradigm, a subset of search items is precued, thereby dissociating the actual set size from the search relevant set size. We hypothesize that a similar process might describe the prioritization that takes place during the search of a realistic scene. Given that the large number of objects in a realistic scene makes an exhaustive item-by-item search strategy impractical, some means must be available to identify a

[1] Even those studies that did manipulate set size in the context of simple scenes (e.g., Zelinsky, 1999; Zelinsky, Rao, Hayhoe, & Ballard, 1997), did so to make the search task more compelling, not to identify scene constraints.

subset of a scene's objects for inspection. We refer to this subset as the *functional set size*.

The objects included in the functional search set likely depend on constraints imposed by the scene. If a person is looking for her friend in a restaurant, she can exclude the wall hangings from her search, thereby restricting the number of patterns requiring inspection. If she suspected that her friend was already seated, she could further restrict her search by focusing only on tables. To do this, perceptual grouping processes (Banks & Prinzmetal, 1976; Grossberg, Mingolla, & Ross, 1994; Treisman, 1982) might be applied to the table features in the scene, thereby selecting these objects for inclusion in the functional search set. Through the enlistment of semantic scene constraints and grouping processes, a large actual set size can be functionally reduced to a more manageable set size.

To explore set size effects in more realistic contexts, and the factors affecting the formation of a functional search set, we had observers search for a target (a military tank) in quasi-realistic landscape scenes. The distractors were trees, whose number and location were controlled. In one condition (Sparse) a small number of trees were scattered randomly over the landscape; in a second condition (Dense) a large number of trees were bunched into groves, creating the appearance of forested areas interspersed with fields. According to traditional conceptions of set size, search should be less efficient in the Dense condition relative to the Sparse condition, reflecting the actual number of distractors in the scene. However, as the number of distractors increase, so does the pressure to enlist scene constraints and grouping processes to create a more manageable functional set size. In this eventuality, one might predict a reverse set size effect, more efficient search in the dense condition relative to the sparse condition.

METHOD

Twelve undergraduates from Stony Brook University indicated the presence or absence of a green tank (approximately 0.65°) in quasirealistic scenes ($27^\circ \times 20^\circ$; 1280×960 pixels), created using Autodesk's 3D Studio Max. The distractors were trees and similar in colour to the target. Each tree subtended approximately $1.5^\circ \times 2.5^\circ$, with sizes varying with scene depth. To avoid target pop-out, the target always appeared near a tree without creating an occlusion. Set size was manipulated in two conditions; 25 trees per scene in the sparse condition (Figure 1a) and 200 trees per scene in the dense condition (Figure 1b). We also manipulated the number of clouds in the sky (0 or 6).

Observers were shown a semblance of the target ($\sim 8.34^\circ$, side view, white background) at the start of the experiment and instructed to make their

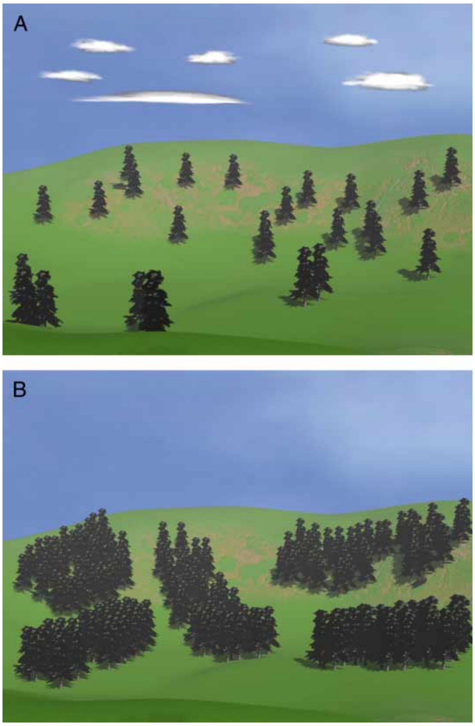

Figure 1. Sample search scenes from the sparse (A) and dense (B) conditions. The target item never appeared closer than 2° to the centre of the display, was never occluded, and was always oriented so as to appear from a side view in the search image. Additionally, the target was often rotated around the y-axis on a trial-by-trial basis (+/− ~0–20°) in order to account for changes in ground slope and to keep scenes looking as realistic as possible. Trees in the sparse condition were placed randomly in the scene. In the dense condition, 40 trees were randomly placed in each of five bounding areas, ranging in size from ~7° to ~10°. The size of each bounding area was maintained across trials, but their locations varied, giving each scene a different look. Position constraints prevented any tree from fully occluding another, but partial occlusions could exist in either condition, with these obviously being more common in the densely forested scenes. To view this figure in colour, please see the online issue of the Journal.

search judgements as quickly as possible while maintaining accuracy. Accuracy feedback was provided. Eye movement and reaction time (RT) data were collected using the EyeLink II, sampling at 500 Hz with an estimated 0.2° spatial resolution. A chinrest was used to minimize head movement, and eye movements were classified into saccades and fixations using the eyetracker's default configuration. Target presence and number of clouds were interleaved over trials; tree density (sparse or dense) was blocked and counterbalanced. There were eight practice trials followed by 80 experimental trials distributed evenly across two blocks.

RESULTS

Manual errors

Adding distractors increased detection accuracy. Misses averaged 14% in the sparse condition and 9% in the dense condition, $F(1, 11) = 5.71$, $p < .05$. Generally, miss rates increase with set size, making this pattern contrary to that typically found in a search experiment. A similar pattern was observed in target absent trials; errors averaged 7% in the sparse condition and 5% in the dense condition, although this difference was not reliable, $F(1, 11) = 0.30$, $p > .10$.

Manual reaction times. RTs for correct trials are plotted in Figure 2 as a function of tree density and target presence. If each tree was serving as a distractor in this task, RTs in the dense condition should have been slower than those in the sparse condition. Instead, we found the opposite pattern, a reverse set size effect. In target present (TP) trials, observers took less time ($\sim$660 ms) to find the target in the dense condition compared to the sparse condition, $F(1, 11) = 17.44$, $p < .005$. A similar ($\sim$1 s), but less reliable trend, $F(1, 11) = 4.4$, $p = .06$, was found in the target absent (TA) data. These findings are counterintuitive, and contrary to those generally reported in the search literature and predicted by search theory. Typically, RTs increase as items are added to a display, due either to these items requiring attentional scrutiny (Treisman & Gelade, 1980), or because they create additional opportunities for confusion with the target (Palmer, 1995). This relationship between set size and search did not hold in our experiment. Although it is

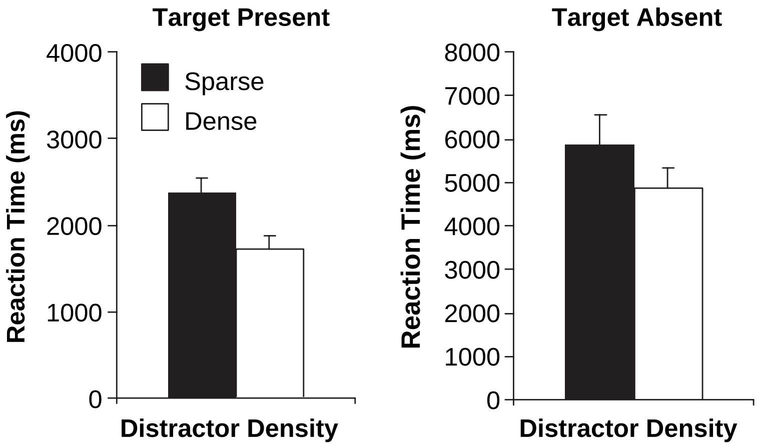

Figure 2. Mean reaction times as a function of distractor density and target presence. Error bars indicate one standard error of the mean (SEM).

possible that more standard set size effects might have emerged under different testing conditions (e.g., if the distractors were not all of one type or were not arranged into clumps), adding tree distractors in our task enabled observers to find the target more quickly.

If set size effects increase with the number of scene objects, we would also expect longer RTs for scenes showing 6 clouds in the sky compared to those showing a clear sky. This was not the case (average difference of 11 ms in TP trials and 48 ms in TA trials; $p > .10$). Although a negative result, this finding supports recent work showing a restriction of search to target-consistent scene regions (Neider & Zelinsky, 2006b; see also Torralba et al., 2006, for a computational model, and Henderson et al., 1999, for early observations). Despite being large and conspicuous objects, observers knew that clouds were confined to the sky, whereas the tank target was confined to the ground, and they used this semantic information to exclude cloud objects from the functional search set.

Analysis of regional gaze preference. The RT analysis tells us that search was more efficient with densely populated scenes, but where were observers looking? To answer this question we analysed the frequency of fixations in tree-covered and open field regions of the scenes. We defined a bounding box around each tree in each image and recorded a fixation as being on a tree if it fell within any of these regions. Fixations not assigned to a tree were recorded as being on an open field. The rare fixations in the sky region of the scene were discarded, as were initial and final fixations in TP trials (so as to avoid position biases).

The results of this analysis are shown in Figure 3 as a function of tree density, normalized by the display area occupied by tree distractors and open fields. Adding trees to the TP displays produced clear and qualitatively different preferences in looking behaviour. As indicated by the crossover interaction in Figure 3, observers preferred to look at the open fields in the dense condition, $F(1, 11) = 6.45$, $p < .05$, but preferred to fixate individual trees in the sparse condition, $F(1, 11) = 40.48$, $p < .001$. No regional fixation preference was found in the dense TA data, $F(1, 11) = 0.69$, $p > .10$, due perhaps to observers adopting a conservative strategy of scrutinizing both field and tree objects before rendering a TA judgement. Gaze was still biased towards individual trees in the sparse TA condition, $F(1, 11) = 30.02$, $p < .001$.

These eye movement data compliment the reverse set size effects observed in errors and manual RTs. When trees were few and sparsely distributed in the scene, observers were more likely to consider them as candidate targets, resulting in more fixations to these objects and longer RTs. However, when there were many trees clustered into dense groups, observers tended to

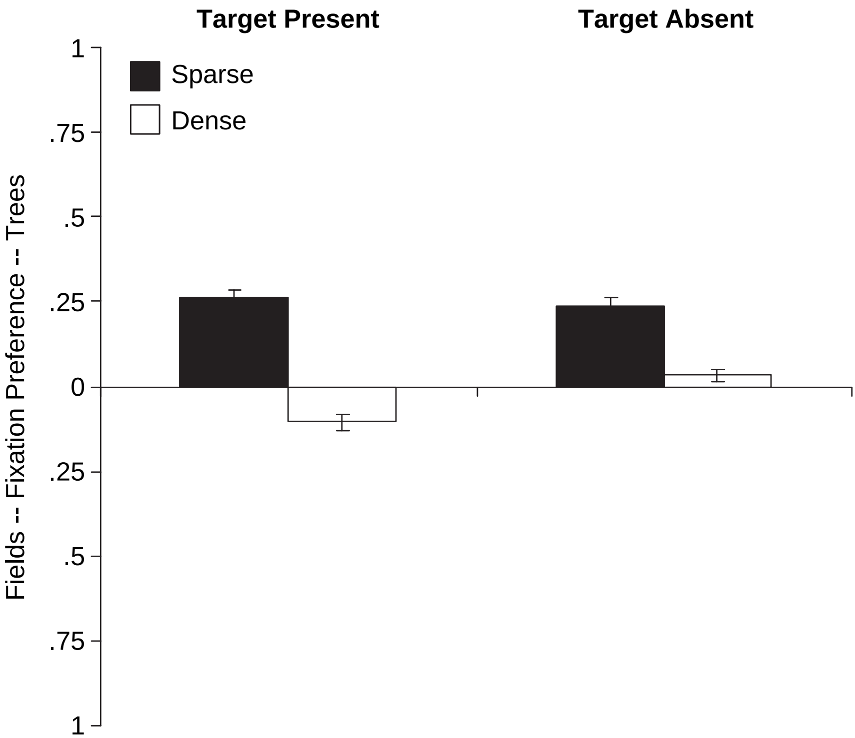

Figure 3. Regional gaze preference normalized by the display area devoted to tree and field distractors. Normalization was accomplished by calculating the total number of pixels in the ground region of the display (all areas below the horizon line; ~657,917 pixels), tree covered regions (~ 115,000 pixels in sparse scenes and ~328,000 pixels in dense scenes), and open regions (~544,000 pixels in sparse scenes and ~328,000 pixels in dense scenes), then dividing the tree and field regions by the total ground pixels to obtain an approximate proportion of each display devoted to each region type. In the sparse displays, trees covered approximately 18% of the ground search area, while open regions accounted for approximately 82% of the area; in dense displays both the tree and open regions occupied roughly 50% of the ground area. After finding the proportions of fixations to the two region types, we then subtracted from these values the respective proportions of each region in the display in order to obtain an unbiased estimate of fixation preference. Using this correction, random fixations in the ground region of the scene would produce no systematic preference for either trees or open fields. Error bars indicate one SEM.

search the open regions of the display, which resulted in faster RTs despite the increase in set size.

GENERAL DISCUSSION

Traditional conceptions of set size cannot be easily extended to search in realistic contexts. Whereas most studies have found increasing RTs with set size, we found the opposite pattern—a reverse set size effect; observers found the target faster as distractors were added to the scene. We also observed qualitatively different patterns of fixation preference at the two set sizes; at low set sizes observers preferred to fixate trees and at high set sizes they preferred to fixate open fields.

We interpret these counterintuitive data as evidence for the creation and use of a functional search set, which determines set size effects in real world search. Search in realistic contexts differs from traditional search tasks in at least two respects. First, there are a staggering number of visual patterns that might be considered objects in even a moderately complex real world scene. This multitude of patterns renders ineffective any simple item-by-item search strategy, and creates the need to reduce the number of candidate search patterns to a more manageable set, what we are calling the functional set size. Second, real world scenes typically depict patterns high in both visual and semantic complexity, and arranged in a meaningful context. We believe that the visual system exploits these characteristics (and likely many others) when extracting the *objects of search* to include in the functional search set.

What were the objects of search in our task? The answer depends on the scene's distractor density. With a small number of distractors sparsely distributed throughout the scene, trees were the objects of search. This is consistent with work showing that objects may be the perceptual units used by attentional and search processes (e.g., Baylis & Driver, 1993; Duncan, 1984; Kramer & Jacobson, 1991; Neider & Zelinsky, 2006a). However, with a larger set size and higher distractor density, observers preferred to fixate the field regions, suggesting that open fields were now the objects of search. This preference for field objects follows if observers learned that the target would never be occluded by trees. We believe that observers used this high-level knowledge to dynamically redefine their objects of search so as to improve their search efficiency, as evidenced by faster RTs in the dense condition. A reverse set size effect resulted from the number of trees in the sparse condition exceeding the number of field objects in the dense condition. As for why observers preferred to inspect individual trees in the sparse condition, we speculate that open fields are less well defined and object-like in sparse displays, leaving only tree objects for inclusion in the functional search set.

Most theories of visual search adopt a relatively static notion of set size; to determine set size one simply need count the number of objects in the display. Extending this logic, these theories might assume that, given sufficiently sophisticated methods of quantifying objects (e.g., Shipley & Kellman, 2001), a similar approach might work for realistic scenes. Our data suggest that efforts to objectively quantify the number of objects in a scene will confront a fundamental limitation, one arising from the largely intangible and highly idiosyncratic factors affecting how observers *choose* to define objects. If an observer adopts one set of criteria over another, new objects will be defined (e.g., trees or open fields) and different patterns of search will emerge. Theories that attempt to attach an exact count to the number of objects in a scene, or to relate search efficiency to the level of clutter in an image, will miss this vital dimension of search and ultimately fail to fully describe search behaviour.

Set size does not exist in a scene; it is not a property that can be objectively quantified, regardless of the sophistication of the counting method. Rather, it is a dynamic interaction between the scene and observer, with the observer playing an active role in what patterns ultimately count as objects of search. She is free to define, and redefine, new objects throughout a search task, with each reconceptualization of the objects of search resulting in different functional search sets and fundamentally different search behaviours.

REFERENCES

Banks, W. P., & Prinzmetal, W. (1976). Configurational effects in visual information processing. *Perception and Psychophysics*, *19*, 361–367.

Baylis, G. C., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 451–470.

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues during search. *Visual Cognition*, *13*, 99–108.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501–517.

Grossberg, S., Mingolla, E., & Ross, W. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review*, *101*, 470–489.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.

Henderson, J. M., Weeks, P., & 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.

Kramer, A. F., & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception and Psychophysics*, *50*, 267–284.

Neider, M. B., & Zelinsky, G. J. (2006a). Searching for camouflaged targets: Effects of target-background similarity on visual search. *Vision Research*, *46*(14), 2217–2235.

Neider, M. B., & Zelinsky, G. J. (2006b). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621.

Palmer, J. (1995). Attention in visual search: Distinguishing four cases of a set-size effect. *Current Direction in Psychological Science*, *4*(4), 118–123.

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). *Feature congestion: A measure of display clutter*. Paper presented at the SIGCHI conference on human factors in computing systems, Portland, OR, USA.

Shipley, T. F., & Kellman, P. J. (Eds.) (2001). *From fragments to objects: Segmentation and grouping in vision.* Oxford, UK: Elsevier Science Publishers.

Toet, A., Kooi, F. L., Bijl, P., & Valeton, J. M. (1998). Visual conspicuity determines human target acquisition performance. *Optical Engineering*, *37*(7), 1969–1975.

Torralba, A., Oliva, A., Castelhano, M., & Henderson, J. M. (2006). Contextual guidance of attention in natural scenes: The role of global features on object search. *Psychological Review*, *113*(4), 766–786.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and objects. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 194–214.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Wolfe, J. M. (1998a). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–71). London: University College London Press.

Wolfe, J. M. (1998b). What can 1 million trials tell us about visual search? *Psychological Science*, *9*, 33–39.

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, *42*, 2985–3004.

Zelinsky, G. (1999). Precueing target location in a variable set size "nonsearch" task: Dissociating search-based and interference-based explanations for set size effects. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 875–903.

Zelinsky, G., Rao, R. P. N., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science*, *8*, 448–453.

Manuscript received February 2007
Manuscript accepted April 2007
First published online July 2007